# ALGORITMI PER IL RESTAURO AUDIO: MODELLI A CONFRONTO

*Sergio Canazza*

Università degli Studi di Udine

*Ma sopra tutte le invenzioni stupende qual eminenza di mente fu quella di colui che s'immaginò di trovar modo di comunicare i suoi più reconditi pensieri a qualsivoglia altra persona benché distante per lunghissimo intervallo di luogo e di tempo? […] Parlare con quelli che non sono ancora nati né saranno se non di qua a mille e dieci mila anni?*

[Galileo Galilei, 1632, *Dialogo sopra i due massimi sistemi del mondo*]

## Introduzione

Negli ultimi dieci anni la ricerca nel campo del restauro audio si è focalizzata sulla progettazione di algoritmi che sottendono una pluralità di modelli e di ipotesi sulla realtà sonora, sviluppati in relazione alla particolare problematica a cui il sistema intende dare risposta:

- Il restauro può essere finalizzato a recuperare l'intelligibilità del parlato durante una trasmissione in ambiente perturbato, dove il fattore critico è il funzionamento in tempo reale, anche a scapito di una forte perdita di qualità del timbro vocale (trasmissioni tra piloti e torre di controllo, tra palombari e nave d'appoggio, tra gruppi militari in terreno nemico);
- Un obiettivo commerciale può invece riguardare la comprensione del parlato durante una trasmissione in ambiente rumoroso: in questo caso si richiede ancora al sistema il funzionamento in tempo reale, ma con una – almeno parziale – preservazione del timbro originale (trasmissione tra dispositivi *mobile* in ambienti quali centri commerciali, feste, concerti);
- Il recupero dell'intelligibilità è anche il fine del restauro in ambiente forense. In questo caso non è richiesto il funzionamento in tempo reale, ma deve essere garantita la fedeltà timbrica per il riconoscimento dell'impronta vocale;
- In questa sede si intendono invece analizzare in dettaglio quegli algoritmi dedicati al restauro di registrazioni musicali, che devono quindi offrire soluzioni soddisfacenti ai problemi connessi al carattere tempo-variante proprio dei segnali musicali. Non è richiesto il tempo reale (spesso il lavoro di un restauratore richiede 10 o 20 volte il tempo reale). Il fine del restauro viene determinato, in questo caso, dai diversi orientamenti al documento adottati, per cui gli interventi di riduzione del rumore potrebbero: 1) rimanere circoscritti ai casi in cui risulti indiscutibile l'evidenza interna della corruttela (approccio documentario) senza mai trascendere il livello tecnologico dell'epoca; 2) essere finalizzati ad un'edizione commerciale (approccio estetico); 3) porsi l'obiettivo di una ricostruzione storica della registrazione così come è stata ascoltata all'epoca (approccio sociologico).

Le metodologie proprie del restauro audio si possono schematizzare in almeno tre diverse categorie, in funzione dell'informazione utilizzata dall'algoritmo durante la fase di attenuazione del rumore.

1) *Metodi in frequenza*; questi algoritmi richiedono poca informazione da parte dell'operatore per realizzare il restauro (informazione *a priori*): è necessario possedere solo una stima del rumore presente (impronta del rumore), che si assume stazionario lungo tutto il segnale. Il resto dell'informazione necessaria (informazione *a posteriori*) viene calcolata automaticamente dal software di restauro analizzando le caratteristiche del segnale. Per la loro semplicità d'uso e per la loro generalizzazione a diverse tipologie di segnali audio, vengono impiegati nei sistemi hardware e software commerciali.
2) Algoritmi nel dominio del tempo che fanno uso di *modelli del segnale*; utilizzano informazione *a priori* per stimare la distribuzione di probabilità degli eventi sonori, il segnale di eccitazione e i coefficienti del filtro. L'algoritmo esegue quindi (informazione *a posteriori*) il tracking del segnale. I modelli generalizzabili a diverse tipologie di segnali sono 'non-informativi' (posseggono poca informazione *a priori*): è necessario quindi particolareggiare il modello di volta in volta, in funzione del segnale in esame.
3) Restauro mediante *analisi per sintesi* (un esempio è illustrato in Sez. 4) e restauro basato su *modelli della sorgente* (proposto da Esquef); in questo caso è richiesta solo informazione *a priori*, rintracciabile attraverso la conoscenza del sistema che ha prodotto il documento audio e l'analisi del materiale sonoro.

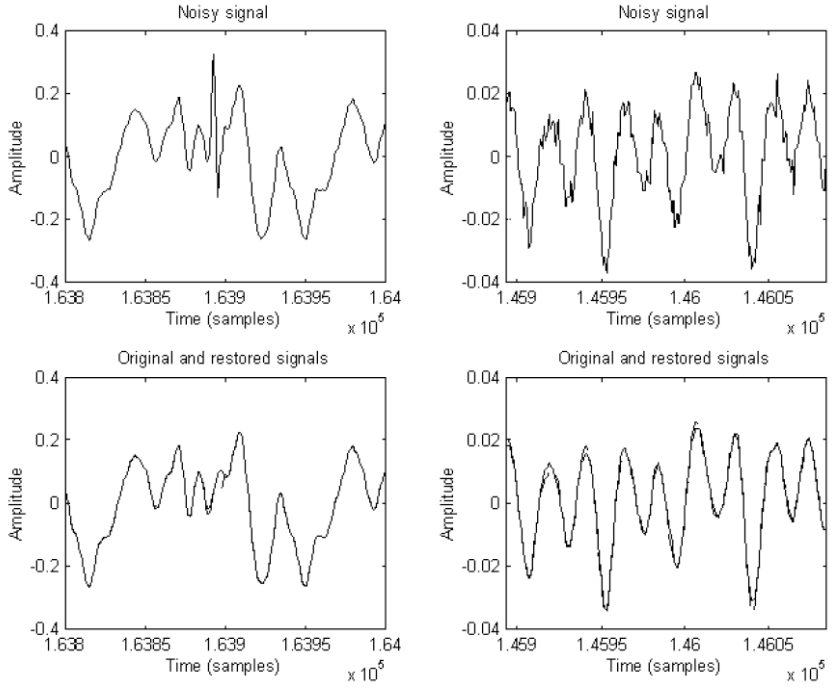



*Fig. 1: segnale con disturbo locale (sinistra in alto) e versione restaurata (sinistra in basso); segnale con rumore globale (destra in alto) e versione restaurata (destra in basso).*

Il degrado a cui va soggetta una registrazione audio è costituito da una serie di disturbi che possono essere sommariamente classificati in due gruppi (v. figura 1):

- disturbi di tipo locale
- disturbi di tipo globale

I *disturbi locali* affliggono solo una porzione del segnale musicale e sono identificabili come discontinuità nella sua forma d'onda. Gli effetti che ne derivano possono variare a seconda della condizione fisica del supporto sul quale si trova l'incisione e dipendono da fattori quali la frequenza del segnale, la durata e la localizzazione del disturbo. A questa categoria appartengono ad esempio i numerosi 'tick' percepibili durante l'ascolto dei comuni dischi in vinile. Questi sono caratterizzati

da una durata breve e sono dovuti principalmente alla granularità del materiale (il vinile) e alla presenza di polvere nei microsolchi. In aggiunta vanno ricordati i graffi superficiali (scratch) e tutti gli altri disturbi (pop, breakage e clipping), caratterizzati da estensioni temporali maggiori. Analoghi artefatti sono presenti anche in altri supporti analogici come le pellicole, i cilindri in cera e i nastri magnetici. È importante sottolineare che anche nelle registrazioni digitali è possibile riscontrare alcuni tick in seguito a perdite di sincronizzazione (jitter) o a errori di scrittura.

I *disturbi di tipo globale* influenzano invece l'intero segnale ed includono il rumore di fondo a larga banda, il wow, il flutter ed alcuni tipi di distorsione non lineare. Il rumore di fondo è forse il tipo di degrado più comune e la sua presenza viene percepita dall'ascoltatore come un fruscio. Questo disturbo viene prodotto dalle parti elettriche e meccaniche delle apparecchiature impiegate per la registrazione, dall'ambiente e dal deterioramento fisico del supporto. In pratica si manifesta in tutte le fasi in cui il segnale è presente in forma analogica. Il suo filtraggio è parte integrante del lavoro svolto negli studi di registrazione e post-produzione e solitamente viene effettuato prima di eseguire la conversione in digitale. Dal punto di vista del restauro, il problema principale risiede nel fatto che tale rumore possiede componenti significative a tutte le frequenze dello spettro audio. Il rischio di andare ad alterare pesantemente la parte utile del segnale è quindi tutt'altro che remoto, e devono essere studiati algoritmi adeguati.

Nella sezione 1 saranno presentati i diversi approcci utilizzati nella rimozione del rumore. In particolare, saranno descritte le metodologie classiche per l'individuazione e la rimozione dei disturbi a carattere locale e globale. Nella sezione 2 saranno illustrati in dettaglio alcuni algoritmi innovativi. Individuare criteri di valutazione degli esiti dei diversi algoritmi di restauro audio è un compito arduo: nella sezione 3 verranno proposti alcuni indici numerici finalizzati a dare un'indicazione della qualità di un restauro musicale e saranno riportati i risultati ottenuti nel restauro di un brano mediante l'impiego degli algoritmi presentati. In sezione 4 viene dato un esempio di restauro mediante *analisi per sintesi*.

# 1. Algoritmi standard

## 1.1. Un approccio basato su modelli del segnale

Le prime tecniche di restauro utilizzate in passato erano analogiche e prevedevano l'editing manuale dei nastri magnetici con una procedura 'taglia ed incolla', allo scopo di rimuovere i difetti locali (ad esempio i graffi), ed una equalizzazione in frequenza per attenuare i degradi di tipo globale, come il fruscio e la distorsione armonica. Poiché i disturbi impulsivi possiedono un elevato contenuto di armoniche ad alta frequenza, si ottennero presto risultati migliori eseguendo un filtraggio passa-alto al fine di rilevarne la presenza e procedendo in un secondo momento alla relativa rimozione mediante filtri passa-basso. Nessuno di questi metodi è però sufficientemente sofisticato da permettere un consistente grado di riduzione dei disturbi senza compromettere in qualche modo la qualità del segnale di partenza.

Il recente sviluppo di tecniche digitali dotate di un più elevato grado di precisione e di maggiore flessibilità operativa consente di raggiungere risultati sensibilmente migliori, ma tali tecniche possono anche provocare sensibili peggioramenti del documento audio, se utilizzate in modo non corretto. Nuove ricerche nel campo del restauro mediante elaborazione numerica si fondano sull'assunzione di modelli matematico-statistici per descrivere la sorgente sonora – modelli della sorgente – o il segnale audio ed i diversi tipi di disturbo che si vogliono considerare – modelli del segnale. In sostanza, si tratta di rigenerare l'audio in funzione del modello utilizzato, la cui scelta, in questo caso, diviene critica. Nel campo della musica elettro-acustica, per esempio, si fa spesso uso di combinazioni di onde sinusoidali modulate in frequenza, con frequenza centrale e ampiezza tempo-varianti. In tal caso è concettualmente ipotizzabile costruire un appropriato modello

generativo di tali segnali modulati in frequenza: l'esatto numero di componenti, le loro frequenze e ampiezze potrebbero essere generate secondo una distribuzione di probabilità opportunamente scelta (in aderenza al modello compositivo usato nell'opera in questione), al fine di sintetizzare segnali artificiali idealmente identici a quelli usati dal compositore e/o dall'esecutore. Queste quantità generate casualmente rappresentano i parametri incogniti del modello, che vanno determinati durante la fase di analisi del segnale esistente.

Si intuisce come la qualità del restauro dipenda in larga misura dalla capacità del modello scelto di descrivere fedelmente il segnale audio. Si supponga di denotare con *x(t)* il valore del segnale musicale campionato all'istante *tT*, dove *T* rappresenta il tempo tra due campioni digitali successivi. Si assume quindi che esista un processo in grado di generare questi campioni digitali che possono essere scritti nella forma $x(t)=f(\theta, e(t))$, dove il vettore $\theta$ rappresenta i parametri del modello, *e(t)* è un rumore casuale, o "eccitazione", termine che descrive qualsiasi elemento casuale presente nel segnale, e *f(.)* è una funzione che permette di passare dalla conoscenza dei parametri e dall'eccitazione al valore dei campioni del segnale audio.

Il termine relativo all'eccitazione *e(t)* può essere interpretato come una deviazione casuale del segnale reale dal modello idealizzato, prima ancora di essere visto come l'effetto di rumore casuale ambientale o comunque estraneo al segnale originale. Il vettore dei parametri $\theta$ potrà essere tempo-variante, per tenere conto di cambiamenti temporali nelle caratteristiche del segnale in considerazione.

In letteratura il modello più usato è quello *auto-regressivo* (AR) a coefficienti costanti di ordine *p*, (modello a soli poli):

$$x(t) = \sum_{i=1}^{p} x(t-i)a_i + e(t) \tag{1}$$

Il valore corrente del segnale a tempo discreto *x(t)*, *t*=1, 2, … è espresso tramite una combinazione lineare dei suoi *p* campioni precedenti a cui si somma il termine di eccitazione *e(t)*. Questo modello è in grado di rappresentare bene sia segnali armonici (poli vicino al cerchio di raggio unitario) che segnali di tipo rumoroso (poli vicino all'origine), e l'ordine *p* può essere interpretato come un indice di complessità della forma d'onda che si vuole trattare.

Poiché le caratteristiche del segnale audio non soddisfano le ipotesi di stazionarietà, se non localmente, va tenuta in considerazione la necessità di aggiornare i coefficienti $a_i$ ad intervalli di tempo regolari (~20ms), ovvero di adottare un modello a coefficienti tempo-varianti (TVAR).

Un altro modello è quello *sinusoidale*, che modella il suono come una somma di sinusoidi con ampiezza $a_k(t)$ e frequenza $f_k(t)$ lentamente variabili nel tempo:

$$\begin{aligned} s_s(t) &= \sum_k a_k(t)\,\cos(\phi_k(t))\ , \\ \phi_k(t) &= 2\pi \int_0^t f_k(\tau)d\tau + \phi_k(0)\ , \end{aligned} \tag{2a}$$

o numericamente

$$\begin{aligned} s_s[n] &= \sum_k a_k[n]\,cos(\phi_k[n])\ , \\ \phi_k[n] &= 2\pi\, f_k[n]T_s + \phi_k[n-1] \end{aligned} \tag{2b}$$

.

In questo modo si riescono anche a descrivere la 'nascita' e la 'morte' di singole componenti frequenziali, in quanto anche il loro numero può variare nel tempo. Lo svantaggio maggiore di tale soluzione consiste nella estrema complessità della fase di identificazione parametrica. Va inoltre tenuto presente che questo modello si presta poco a rappresentare segnali con caratteristiche simili al rumore.

### 1.2. Disturbi locali

Il segnale affetto da disturbi locali può essere modellato diversamente a seconda si ritenga che il click si vada a sommare al segnale musicale (modello additivo) o lo rimpiazzi completamente per un breve periodo (modello sostitutivo). In genere il modello additivo è considerato accettabile per i difetti riscontrabili comunemente, come i tick causati dalla polvere od i piccoli graffi, mentre quello sostitutivo risulta necessario per descrivere i deterioramenti più gravi del supporto di registrazione.
Nel seguito si farà riferimento al seguente modello di validità generale proposto da Godsill:

$$y(t) = x(t) + d(t)v(t) \tag{3}$$

dove *y(t)* rappresenta il segnale disponibile, *x(t)* il segnale non rumoroso, *v(t)* è il singolo click e *d(t)* una funzione indicatrice che assume valore unitario in corrispondenza dei campioni affetti da disturbo e zero altrove. La statistica del processo *d(t)* decide quando un campione è danneggiato mentre quella di *v(t)* determina l'ampiezza caratteristica del disturbo. Chiaramente il valore assunto da *v(t)* è ininfluente quando *d(t)* è nulla, e resta irrilevante anche se *d(t)*=1 nel caso si supponga di procedere alla sostituzione completa dei campioni corrotti.

Anche se esistono diversi approcci per affrontare i rumori di tipo impulsivo, al fine di ottenere la massima qualità tutti i metodi dovrebbero intervenire idealmente solo sui campioni danneggiati lasciando inalterati gli altri, e questo aspetto porta all'individuazione di due fasi distinte nel processo di restauro:

- il *rilevamento* (detection), in cui si cerca di identificare con precisione la posizione e la durata dei disturbi presenti nel segnale utile;
- la *rimozione*, in cui si va a ricostruire il segnale audio ove esso risulti deteriorato, nel modo più accurato possibile.

In realtà, in base a considerazioni psicoacustiche, sarebbe sufficiente rimuovere solo gli artefatti percepibili dall'orecchio umano, ponendosi come obiettivo il raggiungimento del migliore compromesso tra la soppressione dei disturbi e la distorsione inevitabile che si introduce effettuando l'elaborazione.

#### 1.2.1. Fase di rilevamento

Come accennato in precedenza, uno dei metodi più semplici di rilevamento dei click consiste nell'operare un filtraggio di tipo passa-alto sul segnale. Di norma, infatti, al contrario dei disturbi di tipo impulsivo, il comune materiale audio possiede poca informazione per frequenze superiori agli 8 kHz. La fase di detection potrebbe quindi essere realizzata ponendo un rivelatore a soglia in cascata al filtro passa-alto. Questo metodo, che è stato impiegato sia in sistemi analogici che digitali, ha il vantaggio di risultare di facile implementazione e di non richiedere la stima di parametri particolari, a parte il valore della soglia. Alcuni problemi potrebbero insorgere nel caso in cui i disturbi avessero banda limitata o elaborando segnali musicali aventi un ricco contenuto ad alta frequenza.
Un approccio simile può essere utilizzato sfruttando la teoria delle wavelet ed impiegando una tecnica che sostanzialmente equivale al filtraggio tramite un banco di filtri. Per i nostri scopi risultano però più interessanti le metodologie che incorporano alcune informazioni a priori sulle grandezze in gioco, rappresentando il segnale audio mediante un modello AR e sfruttando principi di filtraggio robusto. Per comprendere l'idea alla base di tali tecniche, si supponga che l'andamento del segnale musicale *s(t)* possa essere effettivamente descritto tramite un processo AR localmente stazionario avente in ingresso il rumore bianco *e(t)*. Nell'ipotesi di aver ottenuto dal segnale rumoroso *y(t)* una stima sufficientemente attendibile dei parametri $a_i$ del modello, impiegando allo scopo una qualsiasi procedura di identificazione parametrica robusta ai disturbi impulsivi, dalle equazioni (1) e (3) si può ricavare la seguente espressione per *e(t)*:

$$e(t) = y(t) - \sum_{i=1}^{p} x(t-i)a_i - d(t)v(t) \tag{4}$$

In assenza di disturbi (*d(t)*=0) risulta ovviamente *y(t)*=*s(t)*, e la (4) si semplifica in:

$$e(t) = y(t) - \sum_{i=1}^{p} x(t-i)a_i \tag{5}$$

che evidenzia come applicando ad *y(t)* il 'filtro sbiancante':

$$A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i} \tag{6}$$

si possa ottenere facilmente l'andamento di *e(t)*, il quale in questo caso deve avere le caratteristiche di rumore bianco.

In base a tale osservazione si può pensare, in linea di principio, di individuare i click effettuando un semplice 'test di bianchezza' locale sul segnale *e(t)* in uscita dal filtro (6), impiegando un rivelatore a soglia ed assumendo che ogni irregolarità nell'andamento del rumore possa indicare la presenza di un disturbo. Questo schema permette di amplificare il rapporto tra disturbo e segnale non rumoroso, a scapito però della precisione nella localizzazione temporale del click, che coinvolge ora *p*+1 campioni di *e(t)* in conseguenza alla struttura intrinseca del filtro sbiancante (6). Tale imprecisione può divenire critica in presenza di disturbi ravvicinati e di ampiezza molto diversa, in quanto gli effetti possono sommarsi o cancellarsi reciprocamente. La scelta del valore della soglia, che è influenzato dall'ampiezza dei click, dovrà essere frutto di un compromesso che permetta di individuare il maggior numero di disturbi senza incorrere troppo spesso in falsi allarmi (v. fig. 2).

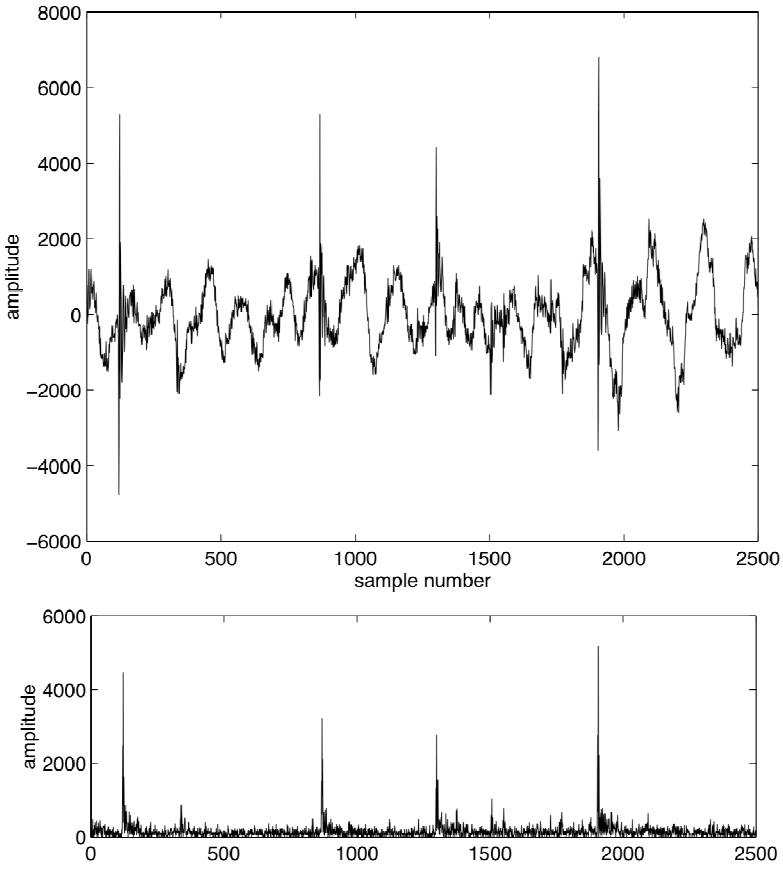



Fig. 2: segnale audio degradato da click (sopra); segnale $e_l(n)^2$ (sotto).

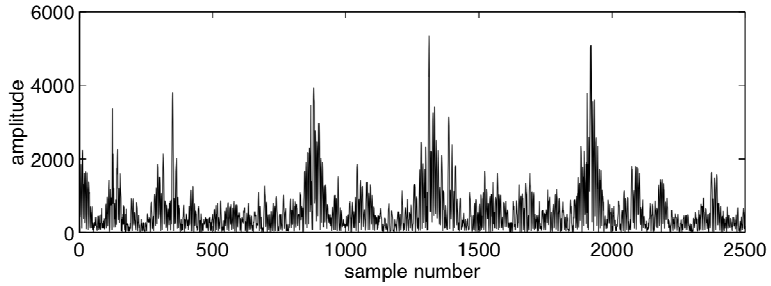


Fig. 3: rilevatore ‘matched filter’.

Esistono anche metodi che impiegano dei filtri ‘matched’ adattati al disturbo impulsivo e, considerando in questo caso il segnale audio come ‘rumore’ colorato additivo, cercano di rilevarne la presenza (v. fig. 3). Altri infine sfruttano le reti neurali.

### 1.2.2. Fase di rimozione

La strategia classica di rimozione dei disturbi impulsivi prevede la ricostruzione completa dei campioni danneggiati mediante tecniche di interpolazione. Dal momento che molto spesso non si può ricavare alcuna informazione utile dal frammento degradato, la sua sostituzione integrale costituisce la scelta più semplice, e va a coincidere con quella adottata in questa sede. Esistono comunque procedure più sofisticate che cercano di estrarre informazioni anche dalla porzione di segnale affetta dal disturbo mediante metodi di modellizzazione del rumore.

Il problema può essere formalizzato nel seguente modo: si considerino $N$ campioni del segnale audio che costituiscono un vettore $x$. Sia $y$ il corrispondente vettore contenente il segnale rumoroso e $d$ quello che indica la presenza dei click. Il vettore $x$ può essere suddiviso in due parti: la prima, che contiene gli elementi il cui valore è noto (cioè in cui $d(t)$=0), verrà denotata con $x_k$; la seconda, relativa ai campioni corrotti e quindi sconosciuti ($d(t)$=1) denotata con $x_u$. In modo analogo si partizionano i vettori $y$ e $d$. Il tutto può essere visto allora come il problema della stima di $x_u$ a partire dai dati osservati $y$.

A questo scopo sono stati sviluppati diversi metodi, il più semplice dei quali è quello del filtro di media. In pratica si sostituiscono gli elementi danneggiati con un valor medio che tiene conto delle caratteristiche della forma d’onda del segnale. Tale approccio non è però adatto a trattare frammenti più lunghi di una decina di campioni.

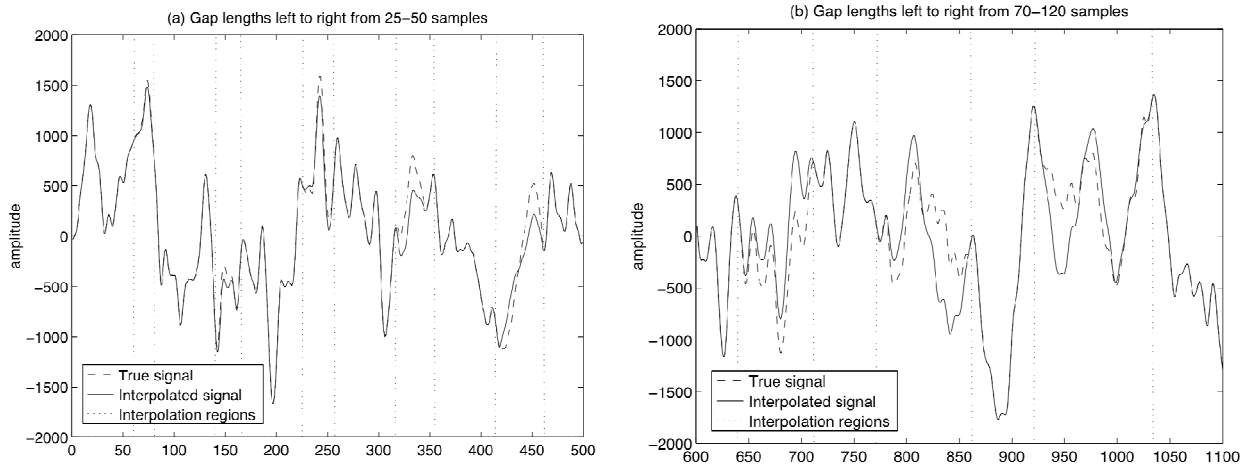


*Fig. 4: interpolazione tramite LSAR, P=60, musica da camera. (a) brevi click (<50 campioni) (b) click prolungati (70÷120 campioni).*

Metodi sicuramente migliori sono quelli che, basandosi sulla definizione di un modello, sfruttano le informazioni dedotte a priori sulle caratteristiche del segnale per stimarne l’andamento a partire dalle misure disponibili. In particolare, nel seguito si accennerà all’interpolatore autoregressivo.

Tale tecnica, denominata *Least Squares AR-based* (LSAR), è stata sviluppata da Vaseghi e Rayner e venne impiegata originariamente per la rimozione dei tick digitali nei CD e DAT (v. fig. 4).

Si consideri un blocco di $N$ campioni che si suppone siano mancanti dal segnale $x$ generato in uscita da un processo (localmente stazionario) di tipo AR avente parametri $a$. Si supponga di aver ricavato una stima dei coefficienti del modello AR a partire dai campioni disponibili prima del frammento danneggiato e subito dopo. Esprimendo $e(t)$ dell'equazione (1) in forma matriciale si ha:

$$e = \boldsymbol{A}\, x \tag{7}$$

in cui $\boldsymbol{A}$ è una matrice di dimensioni $(N\text{-}p)\times N$ in cui la riga $(j\text{-}p)$ è costruita in modo da generare il residuo:

$$e(j) = x(j) - \sum_{i=1}^{p} x(j-i)a_{j,i} \ . \tag{8}$$

Il secondo membro di quest'ultima equazione può essere suddiviso, come illustrato in precedenza, in sezioni comprendenti i dati noti e quelli incogniti, cosicché $\boldsymbol{A}$ risulta partizionata per colonne. La soluzione ai minimi quadrati si ottiene minimizzando l'indice:

$$E = e^T e \tag{9}$$

Come è noto in questo caso la soluzione è esprimibile in una forma matematica chiusa: sono disponibili vari algoritmi che risolvono numericamente il problema. I dati sperimentali dimostrano come l'approccio LSAR fornisca sempre risultati di qualità superiore rispetto agli interpolatori puri (che non fanno assunzioni sulla natura del segnale), e la scelta di ordini elevati per il modello AR può contribuire (entro certi limiti) a ricostruire in modo migliore le porzioni più estese di segnale deteriorato. Questo interpolatore ha la proprietà di essere lo stimatore dei campioni mancanti non polarizzato a minima varianza. Si osservi che un procedimento del genere richiede, comunque, la presenza di una procedura di identificazione parametrica robusta ai disturbi impulsivi , e rappresenta quindi una scelta sub-ottima.

Il problema diventa sensibilmente più complesso se si suppone di avere come incognita anche il vettore $a$ dei coefficienti del modello AR. In questo caso la minimizzazione dell'equazione (8) rispetto a $x_u$ ed $a$, che corrisponde allo stimatore a massima verosimiglianza dei campioni mancanti e dei parametri, contiene fattori indeterminati del quarto ordine e non può essere risolto analiticamente. Una possibile soluzione potrebbe consistere nell'operare prima rispetto ai dati e successivamente rispetto ai parametri: ciò garantirebbe una convergenza almeno locale della funzione di massima verosimiglianza. L'approccio suggerito inizialmente da Niedzwiecki e Cisowski si propone di risolvere il problema in maniera simultanea sfruttando il *Filtro di Kalman Esteso* (EKF).

Si vuole comunque sottolineare che, indipendentemente dalle tecniche impiegate, l'utilizzo di procedure di *overmixing* può migliorare molto la qualità del suono restaurato. Esse consistono nell'effettuare due stime distinte dei campioni danneggiati: la prima elaborando i dati 'in avanti', nella porzione di segnale che precede il click; la seconda è invece da eseguirsi invertendo l'asse temporale (procedendo 'all'indietro') nella zona che segue il disturbo. Il restauro finale verrà poi calcolato mediando le due ricostruzioni in base a un qualche criterio di ottimo od impiegando, ad esempio, come *smoother*, un filtro di Kalman.

Per concludere la trattazione sui rumori impulsivi, si accenna ad una particolare tipologia detta 'impulsiva a bassa frequenza'. Nonostante sia sempre stata trattata come gli altri disturbi, è da sottolineare la sostanziale diversità rispetto ad essi. Essa infatti ha una durata decisamente superiore a quella dei click 'normali' (può arrivare a qualche decimo di secondo) ed è provocata da un difetto marcato, addirittura dalla rottura, del supporto sonoro. Caso tipico nei dischi, la pesante discontinuità induce una oscillazione nella testina, con grande contenuto di energia a bassa frequenza (fig. 5); nei supporti magnetici, questo difetto può generarsi a causa di perdita della pasta magnetica (operazioni di montaggio e trattamento del supporto condotte per mezzo di attrezzature

non opportunamente de-magnetizzate). Altra caratteristica peculiare è il fatto che si ha *sovrapposizione* al segnale audio, e non *sostituzione*. Gli algoritmi progettati per i click non riescono ad eliminare in modo soddisfacente tali difetti. La soluzione innovativa proposta da Godsill prevede una riduzione basata sui *template*. Si sfrutta il fatto che la forma del disturbo è pressoché identica su tutta la superficie del supporto; si memorizzano le caratteristiche di un 'rappresentante' preso in una porzione dove è assente il segnale musicale: sarà quindi sufficiente effettuare una sorta di sottrazione temporale dove si presenta il disturbo aggiunto al segnale utile. Questo metodo non è facilmente implementabile in maniera automatica (è molto difficile la fase rivelazione del disturbo, data la grande energia a bassa frequenza) ed è per questo utilizzato principalmente in modo 'manuale'.

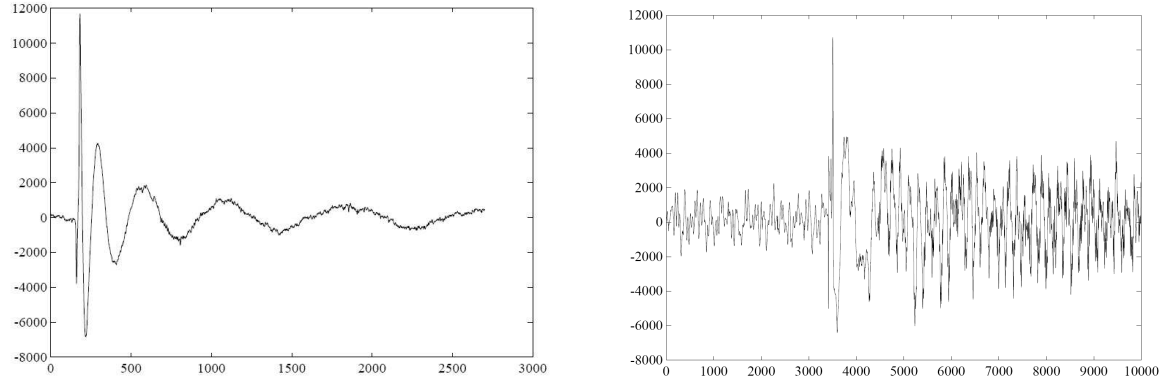



*Fig. 5 Rumore impulsivo a bassa frequenza causato dalle oscillazioni della puntina causate da una rottura in un disco(a). Forma d'onda degradata dal rumore impulsivo additivo(b). Da Godsill 1998..*

### 1.3. Disturbi globali

Il rumore additivo a larga banda (figg. 6 e 7) è la forma più comune di disturbo globale e, possedendo componenti spettrali significative a tutte le frequenze, non risulta in alcun modo eliminabile mediante semplici procedure di equalizzazione. Nella pratica si può spesso ipotizzarne la stazionarietà e, onde evitare problemi di sovrastima, conviene effettuarne il filtraggio solo dopo aver rimosso dal segnale musicale gli eventuali disturbi di tipo impulsivo.

Nel corso degli ultimi decenni sono state sviluppate molte tecniche inerenti alla riduzione del rumore di fondo dalle registrazioni audio. Molte di queste sono accomunate dal fatto di essere basate su un qualche tipo di conoscenza a priori relativa al segnale e/o al rumore.

Una classificazione molto schematica che può essere fatta tra i metodi di restauro esistenti, consiste nella divisione tra gli algoritmi che agiscono nel dominio della frequenza e quelli che invece agiscono nel dominio del tempo. I primi ad essere stati elaborati e i più diffusi sono i metodi nel dominio della frequenza. Generalmente questi non fanno riferimento ad un modello associato al segnale e risultano quindi essere di più agevole messa a punto e di maggior facilità di utilizzo. Di contro i metodi basati sul dominio del tempo fanno spesso riferimento ad un modello del segnale e richiedono quindi una maggior esperienza nell'utilizzo, ma permettono di ottenere, potenzialmente, risultati migliori nella qualità del restauro.

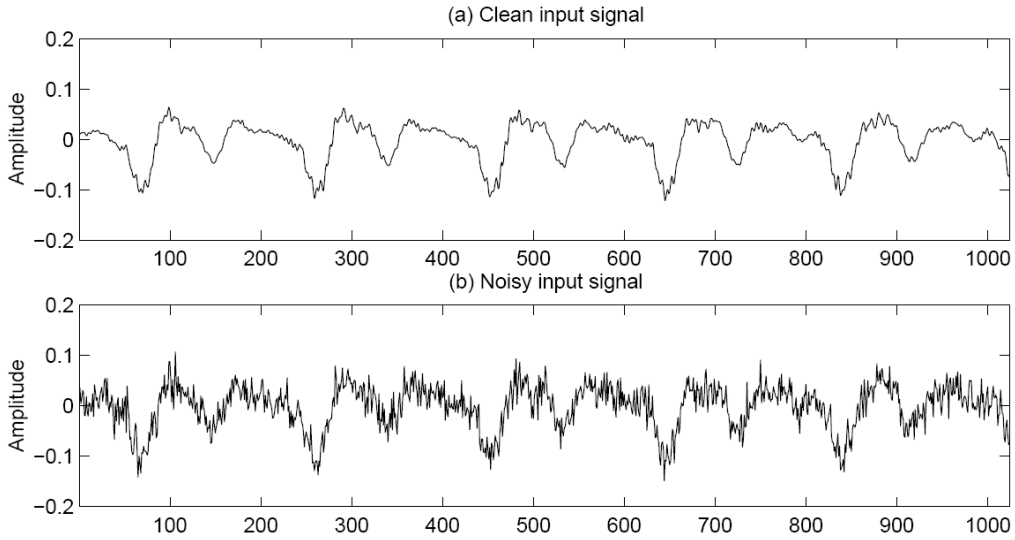



*Fig. 6: esempio di segnale pulito (sopra) e affetto da disturbo globale (sotto). Da Godsill 1998.*

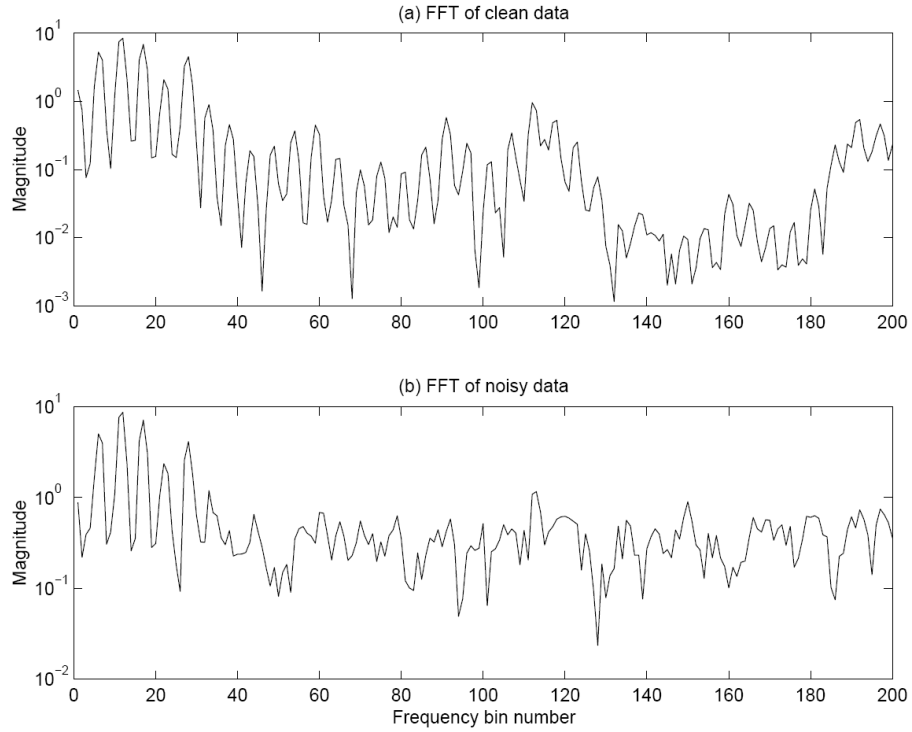



*Fig. 7: spettro dei segnali pulito e affetto da disturbo globale della figura precedente. Godsill 1998.*

### 1.3.1. Metodi nel dominio della frequenza

Le tecniche più diffuse impiegano un'analisi del segnale mediante la trasformata di Fourier a breve termine (cioè calcolata su piccoli tratti di segnale parzialmente sovrapposti: Short-Time Fourier Transform, STFT) e possono essere pensate come un adattamento non stazionario del filtro di Wiener nel dominio della frequenza. In particolare, l'attenuazione spettrale a breve termine (STSA) consiste nell'applicare una soppressione tempo-variante allo spettro *short-time* del rumore e non richiede la definizione di un modello per il segnale audio.

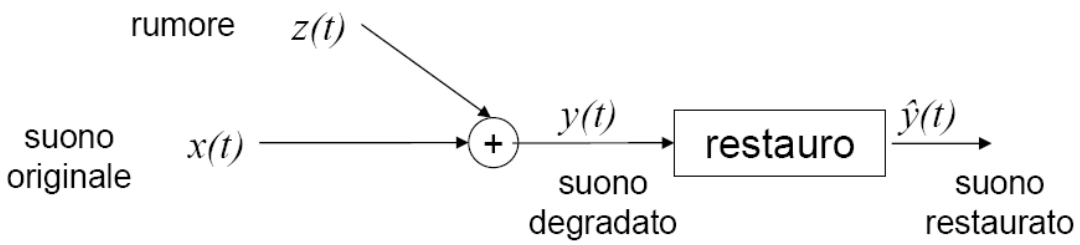



*Fig. 8: modello additivo del rumore a larga banda.*

Si supponga (fig. 8) di considerare il segnale utile *x(t)* come un processo aleatorio stazionario a cui viene sommato del rumore *z(t)* (incorrelato con *x(t)*), per dar luogo al segnale degradato *y(t)*:

$$y(t) = x(t) + z(t) \tag{10}$$

La relazione che lega le rispettive densità spettrali di potenza risulta quindi:

$$P_y(f) = P_x(f) + P_z(f). \tag{11}$$

Nell'ipotesi di riuscire a ricavare durante gli intervalli di silenzio del segnale *y(t)* una adeguata stima di $P_z(f)$, e nei tratti musicali quella di $P_y(f)$, si può pensare di ottenere una stima della densità spettrali di potenza di *x(t)* sottraendo $P_z(f)$ da $P_y(f)$; l'assunzione iniziale di stazionarietà può ritenersi localmente soddisfatta, dal momento che si impiegano finestre temporali di breve lunghezza.

Si noti che l'impiego di un'analisi di tipo *short-time* del segnale è equivalente all'utilizzo di un banco di filtri. Ogni canale (cioè l'uscita di ogni filtro) viene dapprima opportunamente attenuato per procedere poi alla sintesi del segnale restaurato. L'attenuazione tempo-variante applicata ad ogni canale viene calcolata mediante una determinata *regola di soppressione*, la quale si occupa di realizzare una stima (per ogni canale) della potenza del rumore. Ogni particolare tecnica STSA è caratterizzata dal modo in cui viene realizzato il banco di filtri e viene definita la regola di soppressione. Spesso l'analisi a breve termine viene eseguita mediante la STFT. MacAulay e Malpass hanno invece introdotti dei banchi di filtri non lineari. Storicamente, la metodologia STSA è stata sviluppata durante gli anni Settanta del secolo scorso, al fine di rimuovere il rumore nella trasmissione del parlato. Le nuove tecniche STSA per il restauro audio sono un adattamento di queste prime elaborazioni. Tradizionalmente l'interpretazione come STFT è una nozione derivata dall'analisi del parlato.

Un problema aperto rimane quello della fase: nell'interpretazione STFT, l'attenuazione corrisponde ad una modifica del solo modulo dello spettro short-time. È opinione diffusa che la fase non abbia bisogno di essere processata, a causa delle proprietà psicoacustiche dell'orecchio umano. Invero, l''insensibilità alla fase' dell'orecchio umano risulta provata solo nel caso di segnali audio stazionari e per la fase della trasformata di Fourier. Al contrario, nel caso della STFT, variazioni di fase tra i successivi frame short-time può provocare effetti udibili (come una modulazione in frequenza). È importante sottolineare che nelle classiche tecniche STSA non esiste la possibilità di processare la fase, in quanto non si fa nessuna ipotesi sulle caratteristiche del segnale audio.

Sia $Y(p, f_k)$ la STFT del segnale rumoroso *y(t)*, dove *p* rappresenta l'indice temporale di analisi e $f_k = kF_s/N$ la frequenza, con $F_s$ frequenza di campionamento, *N* lunghezza della finestra e $k = 0,\ldots, N-1$. Il risultato dell'applicazione della regola di soppressione può essere interpretato (fig. 9) come l'applicazione di un guadagno $G(p, f_k)$ ad ogni valore $Y(p, f_k)$ della STFT del segnale rumoroso: tale guadagno corrisponde ad una attenuazione del segnale e dovrà quindi essere limitato tra 0 e 1. Il segnale restaurato $\hat{y}(t)$ viene quindi ricavato utilizzando come modulo dello spettro a tempo breve:

$$|\hat{Y}(p, f_k)| = G(p, f_k) \cdot |Y(p, f_k)| \tag{12}$$

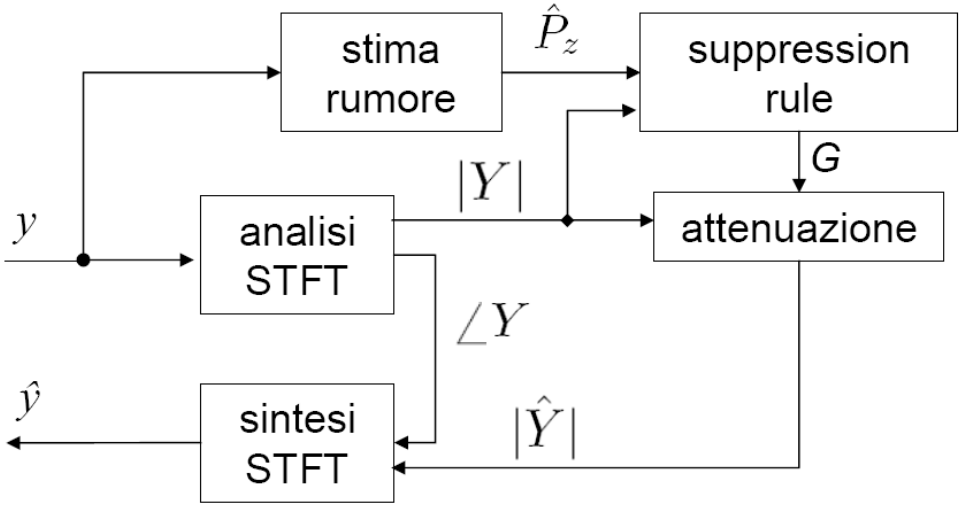



*Fig. 9: Schema del metodo di attenuazione spettrale.*

Nella maggior parte delle regole di soppressione, G(*p, f*$_k$) viene fatto dipendere solo dal livello di potenza del segnale rumoroso (misurata nel medesimo punto) $|Y(p, f_k)|^2$ e sulla stima della potenza del rumore alla frequenza $f_k$,

$$\hat{P}_z(f_k) = E\{|Z(p, f_k)^2|\} \tag{13}$$

(che non dipende dall'indice temporale *p*, a causa dell'ipotizzata stazionarietà del rumore).

È possibile definire a questo punto un segnale *relativo*, cioè il rapporto tra potenza del segnale in ingresso (rumoroso) e potenza del rumore,:

$$Q(p, f_k) = \frac{|Y(p, f_k)|^2}{\hat{P}_z(f_k)} \tag{14}$$

il quale, dall'ipotesi che il rumore *z*(*t*) non sia correlato al segnale *x*(*t*), si deduce essere sempre maggiore di 1.

Una tipica regola di soppressione si basa sul *filtro di Wiener* e si può così formalizzare:

$$G(p, f_k) = \frac{|Y(p, f_k)|^2 - \hat{P}_z(f_k)}{|Y(p, f_k)|^2} = 1 - \frac{1}{Q(p, f_k)} \tag{15}$$

Esistono altre regole, come la sottrazione di potenza (power-subtraction) e la sottrazione spettrale (spectral subtraction), basate sullo stesso principio. Nella figura 10 sono confrontate le caratteristiche di tre regole in relazione al segnale relativo $Q(p, f_k)$. Dal essa si può vedere che le regole di soppressione condividono lo stesso comportamento:

- $G(p, f_k) \rightarrow 1$, quando il segnale relativo è alto ($Q(p, f_k) >> 1$), cioè rumore quasi assente;
- $G(p, f_k) \rightarrow 0$, nel caso sia presente solo rumore (segnale relativo $Q(p, f_k) \rightarrow 1$). In questo senso, in alcuni casi si utilizza una sovrastima della potenza di rumore stimata.

Alcune regole di soppressione, più elaborate, dipendono, oltre che dal segnale relativo, da una conoscenza a priori del segnale corrotto ovvero dalla conoscenza a priori sulla distribuzione di probabilità dei segnali sotto-banda o sul rapporto segnale/rumore. In generale, l'errore commesso da questi procedimenti nel recuperare lo spettro del suono originario ha un effetto udibile in quanto la differenza eseguita tra le densità spettrali può dare, per qualche valore di frequenza, un risultato di segno negativo. Nel caso si decidesse di forzare arbitrariamente a zero i valori risultati negativi si presenterà, nel segnale finale, un disturbo costituito da numerose pseudo-sinusoidi di frequenza

casuale che si accendono e spengono in rapida successione, generando quello che in letteratura è noto come *rumore musicale*.

La procedura di de-noising è illustrata mediante un esempio in cui un segnale musicale (fig. 6a) è corrotto con un rumore bianco additivo gaussiano (fig. 6b). Gli spettri dei due segnali sono mostrati in fig. 7. Si noti come il livello di rumore mascheri le alte frequenze del segnale utile. Sono stati mostrati solo i primi 200 bin di frequenze, corrispondenti a una banda di 8.5 kHz, in quanto c'è poca informazione utile nelle frequenze superiori.

La fig. 11 mostra l'andamento del guadagno *G* del filtro calcolato con i tre metodi di fig. 10: si noti che il metodo di Sottrazione Spettrale opera l'attenuazione maggiore. Questo può essere verificato anche confrontando la forma d'onda dei segnali ricostruiti, dove si nota una chiara progressione del rumore residuo (fig. 12).

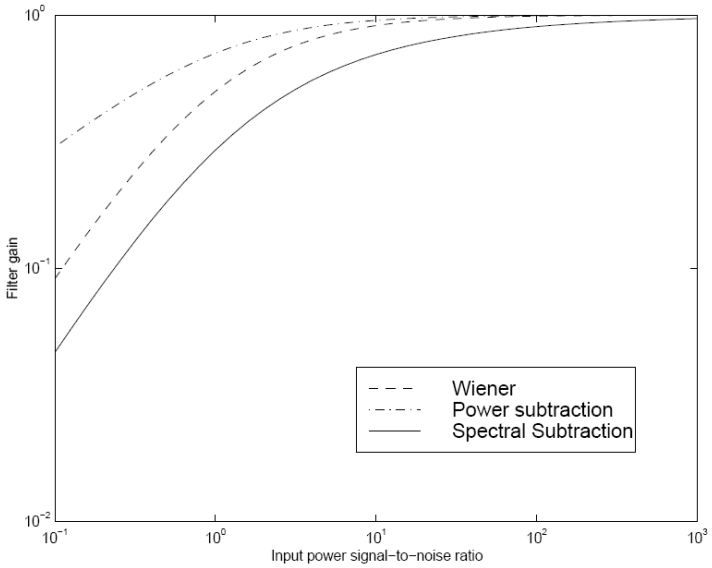



*Fig. 10:* Confronto tra il filtro di Wiener, spectral subtraction e power-subtraction. Godsill 1998.

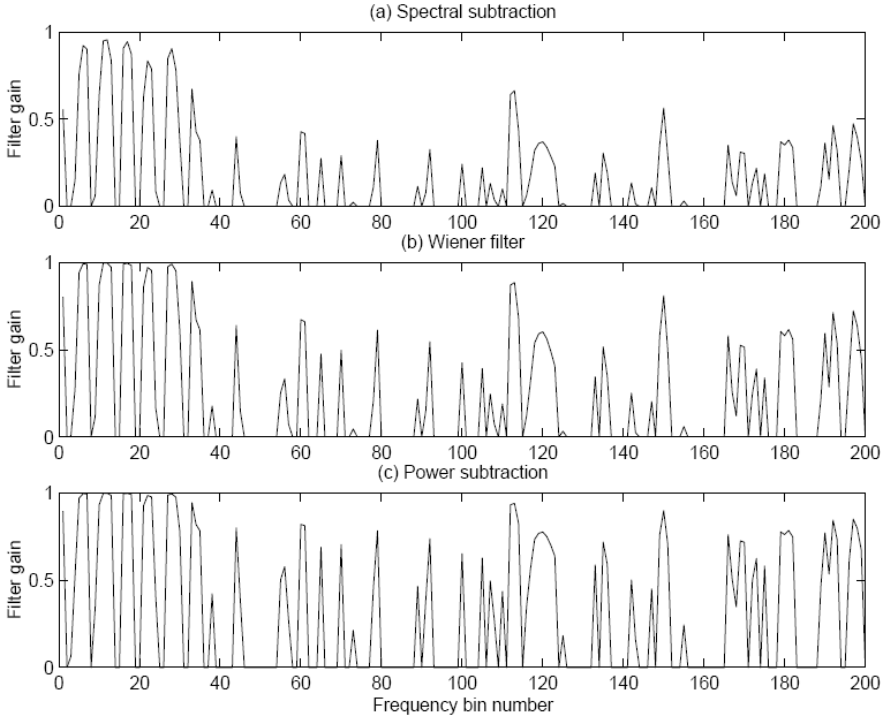



*Fig. 11: guadagno del filtro per (a) sottrazione spettrale, (b) filtro di Wiener, (c) power subtraction. Godsill 1998.*

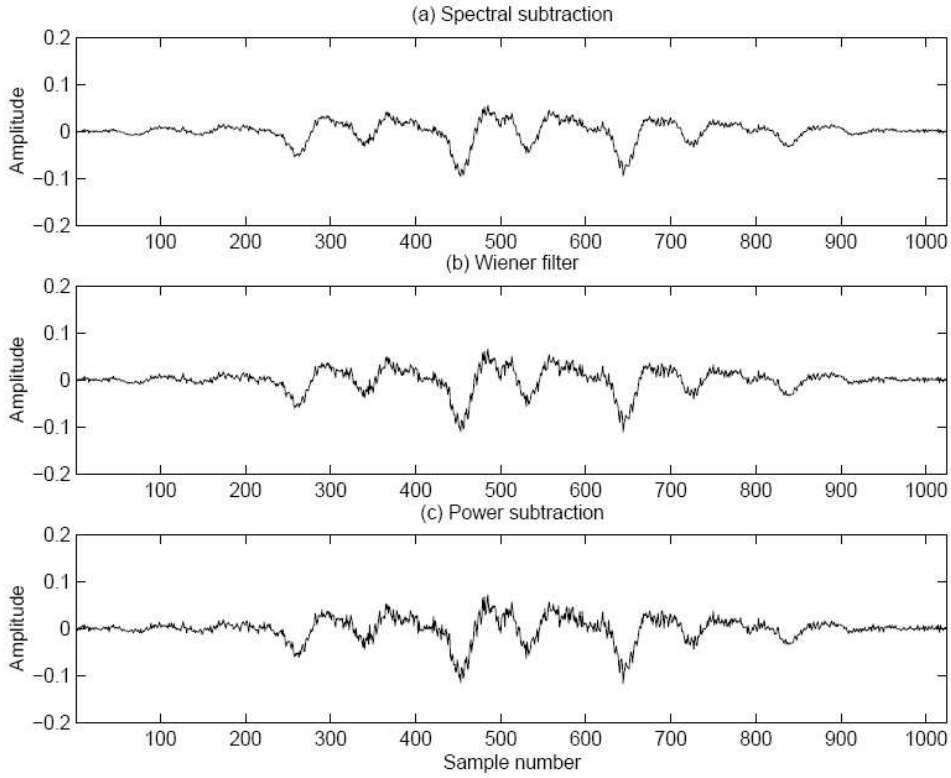



*Fig. 9: confronto tra tre segnali finestrati dopo la riduzione del rumore con i metodi della figura precedente. Godsill 1998.*

In Canazza (2007) sono illustrate nel dettaglio tali regole di soppressione, e sono presentate anche alcune modifiche e innovazioni loro apportate, utilizzando anche considerazioni di tipo psicoacustico relative al fenomeno del mascheramento (*masking*).

### 1.3.2. Metodi nel dominio del tempo

Nonostante i disturbi di tipo globale siano tradizionalmente trattati tramite metodi in frequenza come quelli sopra esposti, recentemente sono stati elaborati metodi che si basano su un approccio nel dominio del tempo e fanno riferimento ad un modello AR per rappresentare il segnale musicale.

Un primo metodo nel dominio del tempo si propone di risolvere in maniera unificata i tre aspetti principali del restauro (identificazione parametrica, rilevamento/rimozione dei click e attenuazione del rumore di fondo) affrontandoli come un problema integrato di filtraggio non lineare per la cui soluzione si può utilizzare una serie di strumenti già consolidati, tra cui anche il filtro di Kalman (KF). Tale schema (v. fig. 13), di impiego molto generale, risolve, in modo ricorsivo e particolarmente adatto a una implementazione numerica, il problema della stima dello stato di un modello lineare. Basato su un meccanismo a due fasi (predizione-aggiornamento) che ne caratterizza le equazioni, esso permette, a seconda del modo in cui viene impiegato, di effettuare operazioni come il tracking dei parametri di un modello, il filtraggio del segnale, o l'interpolazione di campioni mancanti. Il filtro di Kalman gode di alcune importanti proprietà: innanzitutto è lo stimatore lineare che minimizza la varianza dell'errore di stima. Se le distribuzioni di probabilità in gioco sono gaussiane esso diventa anche lo stimatore a minima varianza in senso assoluto. La sua estensione al trattamento di modelli non lineari (denominata Filtro di Kalman Esteso, EKF) effettua, a ogni passo, una linearizzazione 'intelligente' del modello attorno alla miglior stima dello stato disponibile in quel momento, permettendo così di limitare gli errori dovuti a tale operazione. La conseguenza principale dell'impiego di questa approssimazione consiste nella perdita delle garanzie di ottimalità e convergenza.

Un secondo metodo, sempre operante nel dominio del tempo e basato su un modello AR del segnale, utilizza anch'esso le equazioni del filtro di Kalman, ma solo per la fase di attenuazione del rumore di fondo. L'aspetto dell'identificazione dei parametri associati al modello utilizzato viene infatti trattato facendo ricorso a strumenti statistici quali il metodo Monte Carlo. Tale approccio consiste nella generazione di un elevato numero di realizzazioni per il valore dei parametri associati al modello e in una successiva selezione degli stessi in base alla loro probabilità di essere esatti. Una volta ottenuta una stima dei parametri, si fa riferimento al modello del segnale ad essi associato e si applicano le equazioni del filtro di Kalman per ottenere una stima del segnale originale.

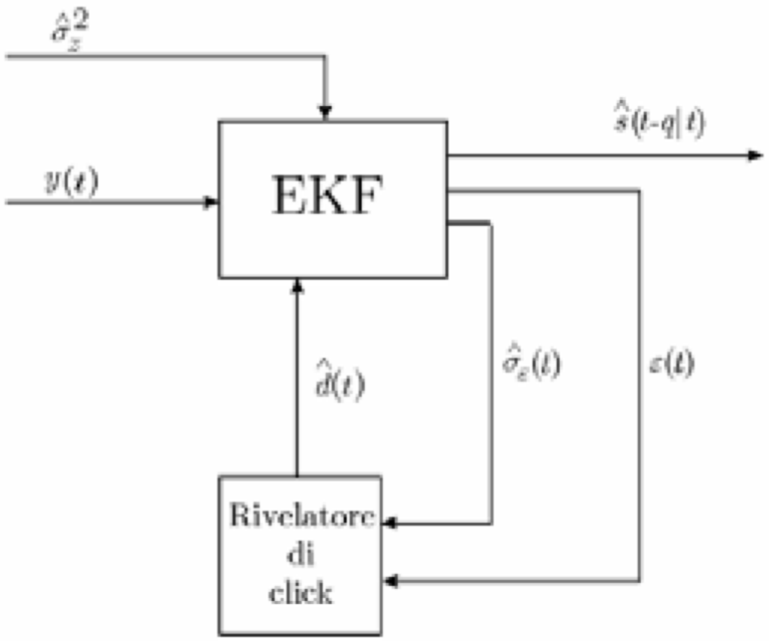



*Fig. 13: schema a blocchi dell'algoritmo EKF.*

# 2. Algoritmi innovativi

## 2.1. Metodi nel dominio della frequenza

### 2.1.1. Filtro di Ephraim-Malah (EMSR)

Successivamente alla soluzione di Wiener sono state proposte molte varianti che comunque soffrono sempre del problema del rumore musicale, anche se in maniera minore. Un sostanziale progresso si è avuto invece con la regole di soppressione spettrale di Ephraim e Malah (*Ephrain Malah Suppression Rule – EMSR*). Questo metodo mira a minimizzare l'errore quadratico medio nella stima delle componenti spettrali del segnale musicale, dei quali $A_k$ e e $\hat{A}_k$: indicano rispettivamente il modulo del *k*-esimo bin frequenziale dello spettro del segnale da restaurare e del segnale restaurato:

$$E\left\{\left(A_k-\hat{A}_k\right)^2\right\} \tag{16}$$

**Dettagli sul metodo di calcolo del guadagno del filtro EMSR**:

Modellando $A_k$ come variabili aleatorie Gaussiane statisticamente indipendenti a media nulla, la soluzione ricavata è la seguente:

$$\hat{A}_k = \Gamma(1.5)\frac{\sqrt{v_k}}{\gamma_k}\exp\left(-\frac{v_k}{2}\right)\left[(1+v_k)I_0\left(\frac{v_k}{2}\right)+v_k I_1\left(\frac{v_k}{2}\right)\right]R_k$$

dove:

$$v_k = \frac{\xi_k}{1+\xi_k}\gamma_k ;$$

$$\gamma_k = \frac{{Y_k}^2}{E\left[\left|Z_k\right|^2\right]} \text{ (}a\ posteriori\text{ SNR)}, \qquad \xi_k = \frac{E\left[\left|X_k\right|^2\right]}{E\left[\left|Z_k\right|^2\right]} \text{ (}a\ priori\text{ SNR)}.$$

e $X_k$, $Z_k$, $Y_k$ sono le componenti spettrali rispettivamente del segnale pulito *x(t)*, del rumore *z(t)* e del segnale rumoroso *y(t)*. $I_0$ e $I_1$ sono le funzioni modificate di Bessel di ordine zero e uno. Si nota che la grandezza $\xi_k$ può essere solo stimata, perché il segnale pulito non è disponibile. Il calcolo della stima viene sviluppato secondo due modelli, uno basato su un approccio a massima verosimiglianza e uno orientato alla decisione (*decision directed*); il migliore si è rivelato il secondo, e perciò riportato di seguito (*n* è l'indice del frame):

$$\hat{\xi}_k(n) = \alpha\frac{\hat{X}_k(n-1)}{Z_k(n-1)} + (1-\alpha)\,P\left[\gamma_k(n)-1\right], \qquad 0 \le \alpha < 1$$

$$P[x] = \begin{cases} x & \text{se } x \ge 0 \\ 0 & \text{altrimenti} \end{cases}$$

In Cappé (1994) viene analizzato il comportamento del filtro basato su tale stimatore; a seguito di un cambio di notazione, il guadagno applicato ad ogni componente spettrale *k* al frame p-esimo è:

$$G(k,p) = \frac{\sqrt{\pi}}{2}\sqrt{\left(\frac{1}{1+Y_{post}(k,p)}\right)\left(\frac{Y_{prio}(k,p)}{1+Y_{prio}(k,p)}\right)}\cdot M\left[\left(1+Y_{post}(k,p)\right)\left(\frac{Y_{prio}(k,p)}{1+Y_{prio}(k,p)}\right)\right]$$

$$M[\vartheta] = \exp\left(-\frac{\vartheta}{2}\right)\left[(1+\vartheta)I_0\left(\frac{\vartheta}{2}\right)+\vartheta I_1\left(\frac{\vartheta}{2}\right)\right]$$

dove i due parametri $Y_{post}$ e $Y_{prio}$ sono calcolati come:

$$Y_{post}(k,p) = \frac{\left|X(k,p)\right|^2}{v(k)} - 1$$

$$Y_{prio}(k,p) = (1-\alpha)P\left[Y_{post}(k,p)\right] + \alpha\frac{\left|G(k,p-1)X(k,p-1)\right|^2}{v(k)}$$

$$P[x] = \begin{cases} x & \text{se } x \ge 0 \\ 0 & \text{altrimenti} \end{cases}$$

dove *v(k)* è la potenza del rumore alla frequenza *k*.

---

In tale metodo il guadagno dipende da due stime diverse del rapporto SNR: $Y_{post}$, che tiene conto dell'informazione del frame precedente, e $Y_{prio}$, corrispondente al tradizionale calcolo di SNR. Il parametro $\alpha$ controlla il bilanciamento tra l'informazione del frame corrente e quella del frame precedente. Giocando su questo parametro si regola l'effetto di smoothing del filtro. Nella figura 14a è possibile notare che $Y_{prio}$ (linea continua) ha minore varianza di $Y_{post}$ (linea tratteggiata). In questo modo si ha minore probabilità che compaia rumore musicale.

Considerando la figura 14b è inoltre possibile vedere l'effetto della variazione di $\alpha$ da 0.98 a 0.998; si nota una stima di $Y_{prio}$ più bassa di circa 10dB ed una varianza decisamente minore. Ciò si manifesta in una riduzione più efficace del rumore musicale. È tuttavia importante notare che al crescere di α si ha un ritardo di risposta sempre più alto nei confronti dei transitori, quindi un effetto passa-basso in occasione di attacchi rapidi del segnale.

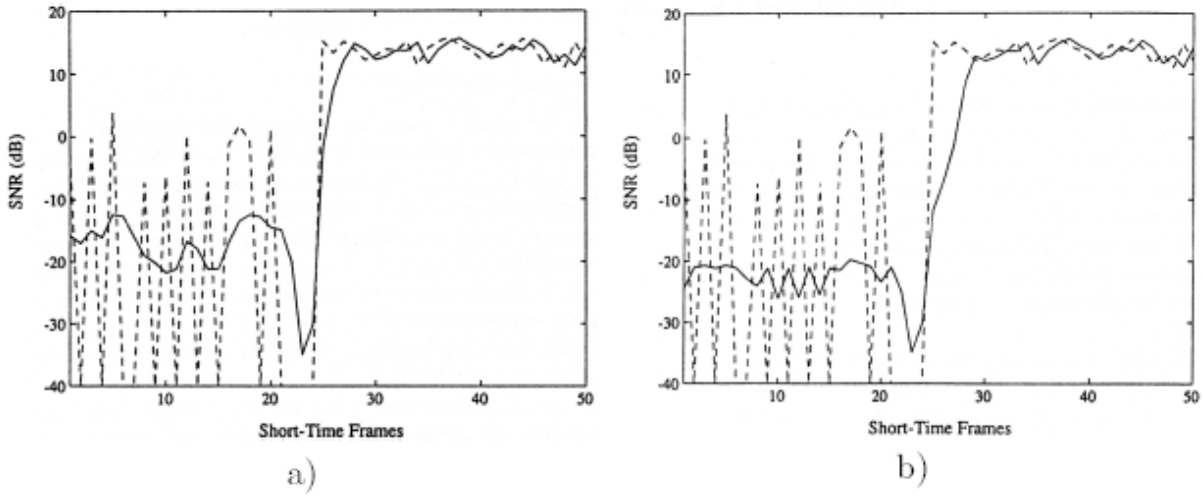



*Fig. 14:* influenza del parametro α (da 0.98 a 0.998).

Un altro vantaggio dell'algoritmo proposto è quello di adattarsi bene al caso di *rumore non stazionario*; ciò è particolarmente importante nel caso di restauro di dischi in vinile, ad esempio 78 giri, dove il rumore si può modellare con ottima approssimazione come ciclostazionario, con periodo uguale a quello di rotazione del disco stesso.

Si deve anche considerare che aumentando la sovrapposizione delle finestre di analisi si aumenta il grado di correlazione statistica tra i frame; ciò si manifesta come una limitazione al potere di riduzione di rumore del filtro. Prove sperimentali hanno portato a concludere che una sovrapposizione spinta (oltre l'80%) può dare risultati accettabili solo aumentando il valore di α.
Una variante di questo metodo (proposta da Mian), chiamata EMSR <alpha>, basata su una stima di $Y_{prio}$ su un numero variabile di frames, ha mostrato un'ulteriore riduzione di rumore musicale residuo.

### 2.1.2. Modello percettivo

I modelli presentati sopra non tengono conto delle caratteristiche dell'orecchio umano. Il modello percettivo qui presentato considera il fenomeno del mascheramento al fine di limitare l'introduzione di artefatti in intervalli frequenziali mascherati dalle componenti del segnale utile. Lo schema generale è riportato nella figura 15.

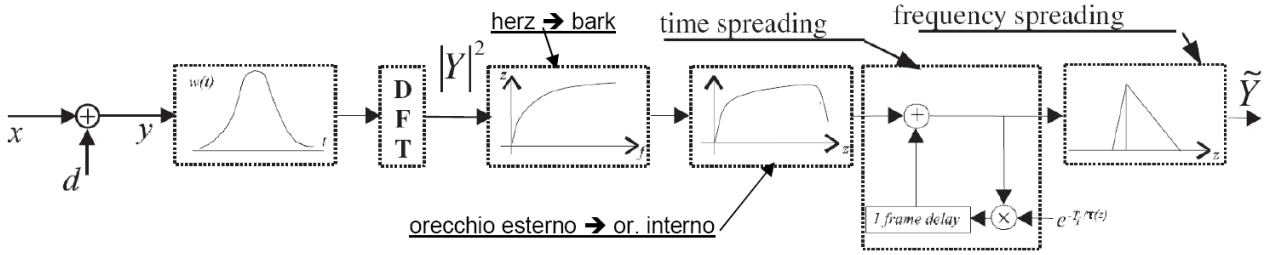



Fig. 15: schema del modello percettivo.

In fig. 15 sono rappresentate, nell'ordine, le seguenti operazioni:

- finestratura nel tempo e DFT con relativo calcolo della potenza del segnale rumoroso ($x(t)$ è il segnale pulito e $d(t)$ è il rumore);
- passaggio dalla scala degli Hertz a quella dei Bark, tramite il calcolo dell'eccitazione del segnale relativo a ciascuna banda critica;
- trasformazione dall'orecchio esterno a quello interno (*outer to inner ear transformation*);

- mascheramento nel tempo (*time spreading*) che è un'operazione con memoria del frame precedente;
- mascheramento in frequenza (*frequency spreading*).

Il segnale $\tilde{Y}$ così ottenuto è la rappresentazione psicoacustica del segnale $y(t)$; le stesse trasformazioni vengono effettuate sull'impronta di rumore disponibile. Con queste rappresentazioni del segnale rumoroso e del rumore viene effettuato un filtraggio con una regola di soppressione analoga a quella di Wiener. Tale soluzione dimostra scarsa attitudine a generare rumore musicale, rispetto al filtro di Wiener nel dominio 'esterno', per merito delle operazioni di spreading che contribuiscono ad attenuare la varianza del filtro, cioè a renderlo ad andamento più dolce e meno frastagliato. Si veda a tal proposito la figura 16 dove sono posti a confronto i guadagni dei filtri psicoacustico e pseudo Wiener (nel dominio esterno).

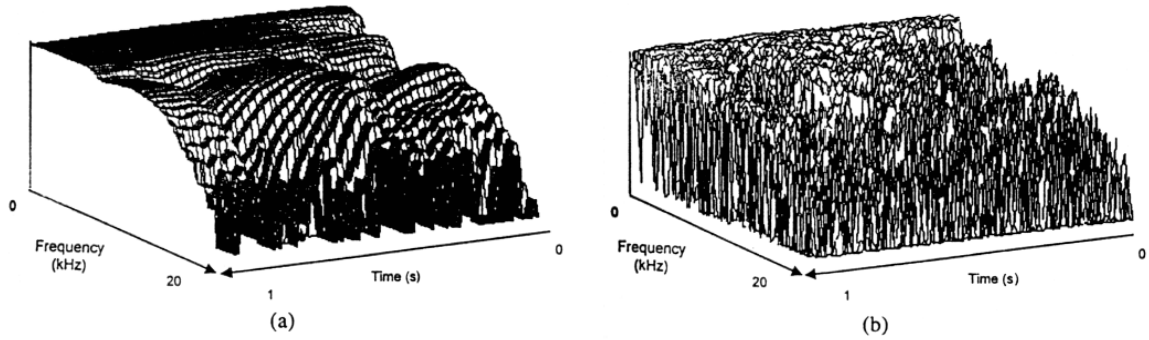



*Fig: 16: confronto tra i guadagni dei filtri psicoacustico e pseudo Wiener.*

Per utilizzare i criteri psicoacustici nella rimozione del rumore è necessario utilizzare una legge di soppressione che consideri l'effetto del mascheramento subito dal rumore stesso. Le soglie di mascheramento che il segnale originale *x(n)* presenta non sono note *a priori* e devono essere stimate. Questa stima può essere ricavata applicando una tecnica standard STSA di riduzione del rumore che porta ad una stima nel dominio della frequenza del segnale originale $\hat{X}(k)$, per il quale le soglie di mascheramento $m_k$, definite come la soglia non negativa sotto la quale un ascoltatore non percepisce rumore additivo, possono essere calcolate utilizzando un opportuno modello psicoacustico.

L'effetto di mascheramento così ricavato viene incorporato in una delle tecniche STSA standard considerando le soglie di mascheramento $m_k$ per ogni frequenza $k$ della trasformata STFT. Viene realizzata una funzione di costo dipendente da $m_k$ che, minimizzata, fornisce la legge di soppressione per la riduzione del rumore. Tale funzione di costo può essere una particolarizzazione del criterio dello scarto quadratico medio per includere le soglie di mascheramento, sotto le quali il costo di un errore è posto uguale a zero.

---

**Dettagli sul metodo del filtro percettivo**

Definendo con $\hat{a}_k$ la stima dell'ampiezza spettrale $a_k$ del segnale originale, la funzione di costo utilizzata è:

$$C(\hat{a}_k, a_k) = \begin{cases} (\hat{a}_k - a_k)^2 - m_k^2 & se \quad |\hat{a}_k - a_k| > m_k \\ 0 & altrimenti \end{cases} \tag{17}$$

La funzione di costo, dato il valore osservato $Y(k)$ e la soglia di mascheramento $m_k$, viene minimizzata calcolando:

$$min_{\hat{a}_k} \int_{a_k} \int_{\alpha_k} C(\hat{a}_k, a_k)\, p(a_k, \alpha_k | Y(k))\, d\alpha_k\, da_k \tag{18}$$

La risoluzione matematica, svolta con metodi numerici, porta ad una legge di soppressione riportata nella figura 17 come funzione del rapporto segnale rumore SNR istantaneo $\varphi(k)$ per diversi livelli di mascheramento. È immediato notare come tale legge di soppressione tenda asintoticamente alla soglia di mascheramento $m_k$ per bassi livelli di rapporto segnale rumore. Inoltre per il caso $m_k = 0$, lo stimatore di ampiezza ritorna ad essere il filtro tradizionale da cui si partiva, nel caso in esame il filtro di Ephraim e Malah.

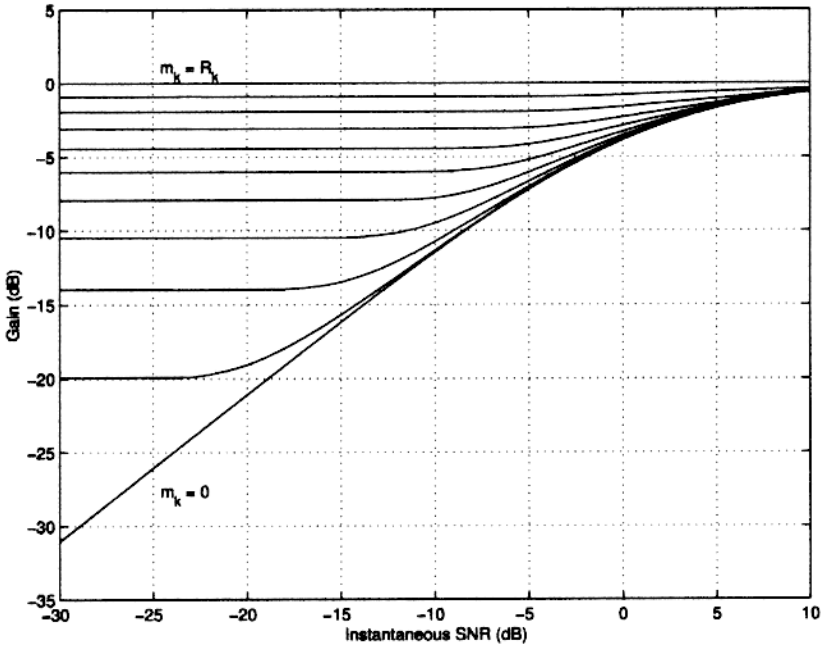



Fig. 17: legge di soppressione in funzione del rapporto segnale rumore SNR istantaneo per diversi livelli di mascheramento ricavata analiticamente.

L'applicazione della regola di soppressione utilizzante il modello percettivo applicato al filtro di Ephraim e Malah richiede numerose e computazionalmente pesanti operazioni matematiche. Per ottenere una discreta efficienza computazionale è preferibile utilizzare una approssimazione della soluzione ottima. È stata utilizzata la legge di soppressione di Ephraim e Malah, calcolata come funzione del rapporto segnale rumore SNR $\varphi(k)$ *a priori*, con un limite inferiore che evita l'attenuazione sotto il livello di mascheramento $m_k$: se l'ampiezza spettrale $Y_k$ del segnale osservato risulta essere minore della soglia di mascheramento, il segnale non viene modificato, dal momento che lo spettro risulta mascherato dal segnale originale.

Anche la stima del segnale originale, utilizzata per calcolare i livelli di mascheramento, viene effettuata utilizzando la legge di soppressione di Ephraim e Malah, intesa come funzione del rapporto segnale rumore SNR *a priori*, ma calcolato utilizzando un valore più elevato del parametro $\alpha$ rispetto a quello utilizzato in EMSR. Tale variazione permette di ottenere una stima del segnale originale con molto meno rumore residuo, ma inevitabilmente con un'attenuazione del segnale molto maggiore. Ciò non è auspicabile se lo scopo è il restauro del segnale, ma è decisamente desiderabile nel processo di calcolo dei livelli di mascheramento associati alla stima delle componenti spettrali del segnale originale, perché consente un calcolo delle soglie tenendo in considerazione solo il segnale originale.

## 3. Validazione dei metodi presentati

In questa sezione sono descritte alcune prove effettuate per testare la validità degli algoritmi presentati nei capitoli precedenti e per valutare il livello di prestazioni raggiunte applicando le implementazioni dei diversi metodi. Nelle numerose sperimentazioni effettuate gli algoritmi sono stati messi alla prova con diversi tipi di generi musicali, con segnali corrotti 'artificialmente' e con registrazioni reali.

### 3.1. Indici numerici per la valutazione “oggettiva” del restauro

Il grave problema che si riscontra nell’impiego di materiale ‘reale’ è rappresentato dalla mancanza di un ‘originale’ cui fare riferimento per le valutazioni qualitative, in quanto le impressioni di ascolto possono variare in modo scarsamente controllabile: esperimenti percettivi hanno mostrato che la valutazione data cambia a seconda se il soggetto è portato a dare maggiore attenzione al segnale musicale oppure al rumore.

Poiché gli algoritmi di restauro lavorano in maniera più efficace nei casi di degrado ‘sintetico’, non si può rinunciare alla possibilità di avere a disposizione un originale non degradato, che permetta di effettuare dei confronti con il segnale restaurato sia a livello percettivo, sia da un punto di vista strettamente numerico: in questa sede si cercherà di dare oggettività alle misure utilizzando alcuni indici di scostamento (rapporti SNR e distanze spettrali). Pur sottolineando infatti che per valutare le prestazioni raggiunte nulla può sostituire l’ascolto diretto, si è deciso comunque di utilizzare indici numerici, al fine di verificare se esistesse una correlazione tra questi e le impressioni percettive e permettere quindi un confronto oggettivo tra i diversi metodi utilizzati.

Nel dominio del tempo una semplice verifica, che può essere fatta avendo a disposizione un segnale pulito $s_{orig}$, consiste nel calcolare il rapporto tra il massimo scostamento tra segnale restaurato $s_{rest}$ e segnale pulito e il valore massimo dello scostamento nel segnale rumoroso (che può essere considerato pari a $3\sigma_e$, nel caso di rumore gaussiano con varianza $\sigma_e^2$) espresso in decibel. È però conveniente utilizzare il valore medio di tale indice calcolato su segmenti della durata di 20ms. Si definisce dunque *massimo assoluto dello scostamento* segmentale:

$$MxD_{\text{seg}} = \frac{1}{M}\sum_{\text{i}=1}^{M} 20\log_{10}\left(\frac{\max|d_{\text{i}}|}{3\sigma_{\text{i}}}\right) \tag{19}$$

in cui si sono indicati con $d_i$ = ($s_{rest}$ - $s_{orig}$) il rumore residuo, cioè la differenza tra il segnale restaurato $s_{rest}$ e l’originale $s_{orig}$ e con $\sigma_i$ la deviazione standard del rumore, calcolati per il segmento *i*-esimo, mentre *M* rappresenta il numero di segmenti utilizzati.

Alcune informazioni utili si possono ricavare anche calcolando il *rapporto segnale-rumore in uscita*:

$$SNR = 10 \cdot log_{10}\left(\frac{\sigma_{orig}^2}{\sigma_d^2}\right) \tag{20}$$

dove $\sigma_{orig}^2$ e $\sigma_d^2$ sono le potenze del segnale originale e del rumore residuo *d* (differenza tra segnale originale e restaurato). Anche per questo indice risulta utile una definizione segmentale:

$$SNR_{seg} = \frac{1}{M}\sum_{i=1}^{M} SNR_i \tag{21}$$

in cui $SNR_i$ rappresenta il rapporto segnale rumore calcolato per il segmento di segnale *i*-esimo di lunghezza pari a 20ms.

Infine si può calcolare anche un indice *SpD* (*distanza spettrale*) che sia legato alle componenti armoniche delle grandezze in gioco:

$$SpD = \int_0^{2\pi}\left(10 \cdot \log_{10}\frac{S_{orig}(\theta)}{S_{rest}(\theta)}\right)^2 \frac{d\theta}{2\pi} \tag{22}$$

In questo caso $S_{orig}$ e $S_{rest}$ rappresentano i periodogrammi del segnale originale $s_{orig}$ e di quello restaurato $s_{rest}$ (calcolati con il metodo di Welch). L’equazione (22) rappresenta una definizione di distanza spettrale spesso impiegata nel campo della codifica del segnale vocale. Infatti essa rappresenta la distanza tra gli spettri espressi in decibel, che essendo una scala logaritmica è più vicina alla percezione umana.

Anche per tale indice è quanto mai utile una definizione segmentale della distanza spettrale che consideri il valore medio calcolato su segmenti di segnale di lunghezza pari a 20ms, dal momento che l'orecchio umano è sensibile ai valori locali più che ai valori complessivi:

$$SpD_{seg} = \frac{1}{M} \sum_{i=1}^{M} SpD_i \tag{23}$$

Uno sviluppo futuro in tal senso potrebbe riguardare la sperimentazione di indici che tengano presenti alcuni aspetti psicoacustici (per esempio l'effetto di mascheramento o le curve di risposta media dell'orecchio umano). Si evidenzia comunque la scarsa attitudine a descrivere realmente la qualità sonora manifestata dalle misure 'strumentali classiche' come quelle appena introdotte.

Il tentativo di scegliere la modalità di filtraggio o i parametri relativi al filtro solo in funzione degli indici numerici non sempre si rivela un metodo di scelta ottimale. Molto utile per esprimere un giudizio sull'efficacia dell'operazione di restauro può rivelarsi l'ascolto diretto della differenza tra il segnale restaurato e il segnale originale. Si riesce in tal modo a verificare se si sia eliminata solo la componente rumorosa o anche qualche componente musicale.

Nei casi reali in cui non è possibile avere il segnale pulito si ricorre a misure soggettive spesso utilizzando le metodologie di test definite dal gruppo MPEG, per valutare la qualità audio della codifica digitale, come lo standard MUSHRA.

### 3.2. Validazione dei metodi nel dominio della frequenza

Essendo molto numerosi i parametri su cui è possibile agire, è stato scelto un sottoinsieme rappresentativo di essi; in particolare, si è scelto di partire dalle versioni rumorose con rapporto segnale/rumore più alto ed effettuare una prova con ogni filtro implementato, lasciando invariati (usando i valori di default) tutti gli altri parametri Così facendo si è cercato di ottimizzare, per quanto possibile, il funzionamento di ciascun filtro. Gli esiti di tali prove possono essere ascoltate in: http://smc.dei.unipd.it/restoration.html.

Dalle prove condotte appare evidente come nell'utilizzo del filtro di Wiener e pseudo-Wiener si abbia una notevole quantità di rumore residuo modulato intorno alle frequenze del segnale: si nota una forte presenza di rumore musicale. Il filtro EMSR presenta un rumore musicale decisamente minore (percettivamente non rilevante per alti SNR), ma è necessario regolare il parametro $\alpha$ al fine di trovare un giusto compromesso tra l'attenuazione del rumore e la distorsione dei transienti.

Il filtro basato sul modello psicoacustico funziona bene con segnali con basso SNR e quindi molto disturbati, mentre per i segnali con elevato SNR il restauro ottenuto a volte risulta particolarmente pervasivo ed elimina molte componenti, soprattutto alle alte frequenze, del segnale originale, come è possibile notare ascoltando la differenza tra il segnale rumoroso in ingresso e il segnale restaurato in uscita.

---

**Dettagli confronto metodi di restauro**

Dalle prove fatte si ricavano delle importanti indicazioni: si è scelto di variare i due parametri fondamentali della dimensione e della sovrapposizione delle finestre di analisi. Come discusso in Canazza (2007), una elevata sovrapposizione, oltre l'80%, non porta un vantaggio nella riduzione del rumore: è invece evidente il vantaggio dell'aumento della dimensione della finestra, a patto di far crescere anche la sovrapposizione (per rendere costante il passo di avanzamento in termini di numero di campioni). Tuttavia, una finestra più grande presenta il difetto di lasciare una maggior quantità di rumore residuo. Si è concluso che le dimensioni di 4096 e 8192 campioni rappresentano un buon compromesso (si considera una frequenza di campionamento di 44.1 kHz).

In presenza di un'elevata quantità di rumore rispetto al segnale (SNR < 15dB) sorge la necessità di sovrastimare l'impronta di rumore per effettuare una riduzione efficace. In particolare, il rumore musicale può essere 'colorato' agendo sulla sovrastima delle alte frequenze: modificando gli ultimi Bark dell'impronta di rumore, è possibile rendere meno udibile tale difetto.

Bisogna sottolineare che a volte, principalmente nei brani fortemente degradati, si possono ottenere risultati migliori operando due successivi filtraggi. Infatti utilizzando un unico filtraggio risulta necessario ricorrere ad una sovrastima dell'impronta di rumore per evitare che il segnale risultante conservi ancora una percettivamente apprezzabile quantità di rumore; questa sovrastima comporta però un filtraggio troppo spinto che modifica anche il segnale utile, condizione che dovrebbe essere evitata, per quanto possibile, in un buon restauro. Operando con due fasi successive è invece possibile ottenere un restauro in grado di preservare il segnale utile. I dati riportati nella tabella 1 sono relativi ad un singolo utilizzo dei filtri.

Un ulteriore parametro regolabile relativo al modello psicoacustico utilizzato consiste nel numero di bande in cui viene diviso lo spettro (25, 50 oppure 100 bande). All'ascolto si nota che il modello psicoacustico implementato funziona decisamente meglio con una divisione che utilizza un basso numero di bande, poiché utilizzando un numero elevato di bande si ha una notevole quantità di rumore residuo che può essere attenuato solo aumentando la sovrastima dell'impronta di rumore, operazione che comporta un evidente degrado del segnale originale. Si è dunque preferito utilizzare la divisione dello spettro in sole 25 bande per ogni restauro. L'ultima scelta consiste nell'utilizzo del modello psicoacustico di Kokkinakis o di Beerends. Nei risultati ottenuti utilizzando i due diversi modelli non si evidenziano differenze.

| | MxDseg | SNR | SNRseg | SpD | SpDseg |
|---|---|---|---|---|---|
| **Rumoroso 10db** | / | 10.00 dB | 4.69 dB | 575.33 dB | 381.28 dB |
| **Wiener** | -5.23 dB | 16.29 dB | 11.16 dB | 433.00 dB | 259.94 dB |
| **Pseudo-Wiener** | -1.96 dB | 13.04 dB | 7.81 dB | 509.05 dB | 324.11 dB |
| **EMSR** | -5.69 dB | 46.91 dB | 11.74 dB | 411.95 dB | 249.24 dB |
| **Psicoacustico** | -6.77 dB | 16.49 dB | 12.84 dB | 88.21 dB | 38.39 dB |

| | MxDseg | SNR | SNRseg | SpD | SpDseg |
|---|---|---|---|---|---|
| **Rumoroso 20db** | / | 20.00 dB | 14.61 dB | 396.6 dB | 238.37 dB |
| **Wiener** | -4.13 dB | 25.15 dB | 19.98 dB | 280.83 dB | 148.45 dB |
| **Pseudo-Wiener** | -1.6 dB | 22.63 dB | 17.36 dB | 342.34 dB | 195.39 dB |
| **EMSR** | -4.08 dB | 25.33 dB | 20.17 dB | 264.00 dB | 140.62 dB |
| **Psicoacustico** | -3.04 dB | 23.54 dB | 19.47 dB | 90.24 dB | 36.39 dB |

| | MxDseg | SNR | SNRseg | SpD | SpDseg |
|---|---|---|---|---|---|
| **Rumoroso 30db** | / | 30.00 dB | 24.54 dB | 251.82 dB | 132.56 dB |
| **Wiener** | -1.86 dB | 32.81 dB | 27.84 dB | 193.13 dB | 92.31 dB |
| **Pseudo-Wiener** | -1.09 dB | 32.04 dB | 27.01 dB | 209.17 dB | 102.78 dB |
| **EMSR** | -3.11 dB | 34.10 dB | 29.30 dB | 149.35 dB | 66.59 dB |
| **Psicoacustico** | -1.04 dB | 32.29 dB | 27.77 dB | 65.32 dB | 21.45 dB |

*Tabella 1. Valore dei diversi indici di qualità in funzione dei restauri ottenuti utilizzando i diversi filtri nel dominio della frequenza.*

Nella tabella 1 vengono riportati gli indici numerici discussi sopra associati ai restauri ottenuti utilizzando i diversi filtri nel dominio della frequenza, di volta in volta ottimizzati nella scelta dei valori da attribuire a ciascun parametro. Il segnale musicale consiste in un frammento (6 sec) di un brano per sola voce cantante (*Tom's Diner* di Suzanne Vega, dal CD A&M, 5136-2) precedentemente corrotto sommando al segnale originale rumore gaussiano bianco con diversa intensità e quindi con diversi livelli di SNR (rispettivamente: 10 dB, 20 dB e 30 dB).

Nella figura 18 è riportato l'andamento del guadagno introdotto da ciascun filtro al variare dell'SNR del segnale rumoroso nel restauro del brano sopra citato. Con guadagno si intende la differenza tra l'SNR del segnale restaurato e l'SNR del segnale di ingresso. Va comunque sottolineato che il valore dell'SNR è solo indicativo della qualità del restauro: ad esempio, nel filtro psicoacustico, che sfrutta il mascheramento del rumore da parte del segnale utile, la componente rumorosa residua – ma mascherata – contribuisce ad abbassare tale indice pur non essendo percepibile. Si evince comunque il buon comportamento, per tutti gli SNR d'ingresso, del filtro EMSR <alpha>.

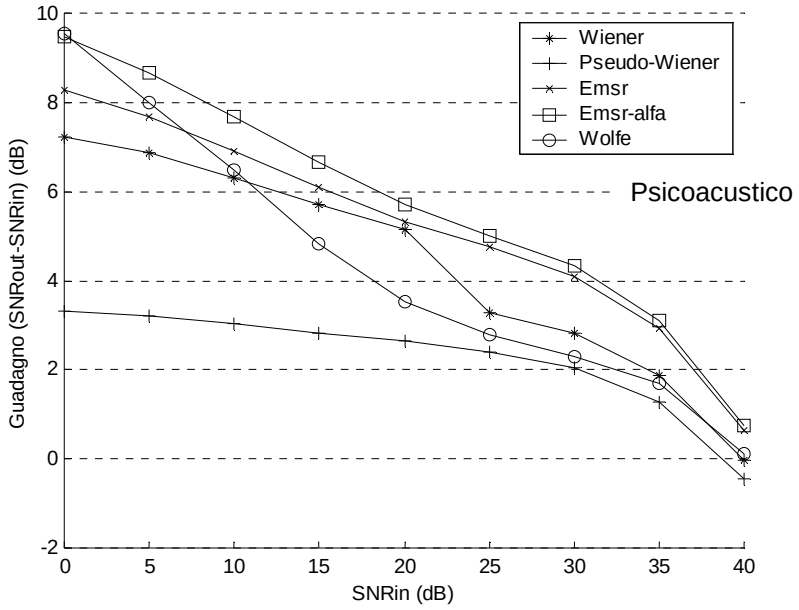



*Fig. 18: andamento del guadagno introdotto dai filtri nel dominio della frequenza al variare dell'SNR di ingresso. Si può notare il miglior guadagno del filtro EMSR <alpha> per tutti gli SNR.*

## 4. Restauro mediante analisi per sintesi

Un approccio (particolarmente adatto alla musica elettroacustica) diverso rispetto ai metodi nei domini della frequenza e del tempo, prevede lo studio del modello compositivo e della sua realizzazione analogica, al fine di separare il *segnale utile* dal *rumore*, inteso nel duplice senso di segnale non-musicale (rumore a larga banda o impulsivo) e di scostamento dal modello proposto dal compositore. In questo modo si possono studiare le due tipologie di rumore, sia a fini restaurativi (maggiore è l'informazione sul rumore, più precisi possono essere i parametri dei modelli utilizzati nel restauro), che a fini storico-musicologici (il rumore può rivelare informazioni sull'equipaggiamento usato dal laboratorio dove è stata prodotta l'opera).

Nel campo della musica elettroacustica, alcune composizioni sono costituite da combinazioni di segnali audio elementari (sinusoidi) opportunamente modulati in ampiezza e in frequenza attraverso indici tempo-varianti. Se sono noti il modello matematico utilizzato dal compositore e la tecnologia adottata nella realizzazione dell'opera, è ipotizzabile ricostruire il segnale musicale sintetizzando le diverse componenti frequenziali, secondo il metodo di "analisi per sintesi". Questo prevede i seguenti passi:

a) misura dei parametri degli oggetti audio: durata, spettro, intensità, inviluppo. Le variabili da studiare dipendono dalle ipotesi di lavoro e dalle possibilità tecniche degli strumenti di misura;
b) selezione e analisi delle variabili più importanti al fine di caratterizzare l'opera, in funzione del modello matematico adottato dal compositore;
c) interpretazione numerica dei dati: scelta delle diverse rappresentazioni e scale temporali;
d) sintesi dell'opera musicale, in accordo con i dati emersi dalle misure;
e) studio della relazione tra l'opera originale e quella sintetizzata, al fine di focalizzare correlazioni tra diversi parametri, individuare le variabili più importanti e determinare eventuali variabili da eliminare;
f) ripetizione della procedura (passi b÷e) sino a che il risultato converge.

## Esempio di restauro mediante sintesi: *Analogique B*

A titolo d'esempio, si applica questa metodologia a *Analogique B* di Iannis Xenakis. Xenakis descrive i procedimenti tecnici e la struttura matematica dell'opera, dai quali risulta comunque impossibile riprodurre l'opera. Si è quindi applicata la strategia di analisi per sintesi, modellizzando la composizione direttamente dall'analisi del materiale sonoro. La composizione, per nastro magnetico, realizzata nel 1959 presso gli studi del GRM (Groupe de Recherches Musicales) di Parigi, è basata su un modello di sintesi granulare ispirato dalla teorie di Gabor. A titolo esemplificativo, si è preso in esame un estratto (6 s) preso da un esemplare fornito da Ricordi-Salabert.

Xenakis ha descritto in questo modo il modello di sintesi granulare per Analogique B:

all sound, even all continuous sonic variation, is conceived as an assemblage of a large number of elementary grains adequately disposed in time. […] In the attack, body, and decline of a complex sound, thousands of pure sounds appear in a more or less short interval of time, $\Delta t$.

Xenakis ha pensato di organizzare i grani sonori in *screen*, una sorta di intervalli temporali all'interno dei quali vengono distribuiti in maniera statistica centinaia di grani. Risulta che il compositore abbia utilizzato esclusivamente sinusoidi, generate attraverso l'uso di oscillatori analogici. La durata dei grani, in accordo con l'intenzione dell'autore – «cogliere lo spessore del presente» – è riconducibile a 40 ms, ma non esistono elementi certi per escludere valori diversi. Per quanto concerne l'intensità, e quindi l'ampiezza dei grani, Xenakis scriveva che avrebbe usato solo quattro livelli. Un grano sonoro è dotato di una propria frequenza, un'ampiezza, una durata e un inviluppo. La durata minima necessaria per distinguere il pitch dipende dalla frequenza del segnale: è necessaria una soglia di 13 ms per le alte frequenze, che cresce fino a 45 ms per le basse. Al di sotto di tali limiti il grano è percepito come un *click*.

Il lavoro si è articolato nelle seguenti fasi: determinazione delle frequenze, dei tempi e delle ampiezze di ciascun grano, calcolo dei parametri dei filtri per l'individuazione degli inviluppi, risintesi del brano utilizzando i risultati dell'analisi, al fine di evidenziare, per sottrazione con la registrazione storica, alcuni aspetti del rumore imputabili alla tecnologia dell'epoca, e quindi aiutare a modellare il comportamento delle apparecchiature storiche.

Per la discriminazione delle frequenze, si è utilizzata l'autocorrelazione al fine di discriminare le frequenze delle sinusoidi che costituiscono i grani. Da un'analisi effettuata sull'intero segnale si è osservato che oltre i 10 kHz non compaiono grani; si sono quindi considerate frequenze solo fino a questo valore. Inoltre suoni di ampiezza inferiore a -96 dB sono stati considerati rumore.

Per ricostruire gli inviluppi, sono state utilizzate due diversi tipologie di filtri in cascata: passa banda, centrati su ciascun valore emerso dall'analisi, al fine di ottenere i grani relativi alle rispettive frequenze; passa basso, per ottenere l'inviluppo dei grani. È stato necessario limitare la selettività dei filtri per ottenere un tempo di risposta compatibile con la durata dei grani, al fine di ottenere una valutazione corretta dell'ampiezza del segnale.

Per valutare correttamente l'inviluppo del segnale, si sono seguite due modalità differenti in base alla frequenza del segnale: interpolazione dei picchi massimi e filtraggio passa basso. La frequenza di 'soglia' tra le due modalità è stata posta a 140 Hz: al di sotto di tale frequenza, il filtraggio non era in grado di eliminare la frequenza audio del grano e quindi non forniva correttamente l'inviluppo. L'analisi ha permesso di ottenere i campioni d'inizio e di fine dei grani e la loro ampiezza.

Gli inviluppi stimati sono ritardati – circa 7.25 ms (320 campioni) – rispetto a quelli reali e il segnale risulta scalato proporzionalmente in ampiezza. Si sono dunque effettuate alcune correzioni:

1) gli inviluppi ricostruiti per interpolazione dei picchi, realizzati solo per frequenze inferiori ai 140 Hz, sono stati ritardati di 320 campioni in modo da conservare la corretta temporizzazione relativa rispetto a quelli ottenuti per filtraggio;
2) gli inviluppi ottenuti per filtraggio sono stati amplificati di un valore pari al rapporto tra il massimo del segnale prima del filtraggio passa basso e il massimo dell'inviluppo, ripristinando così i reali valori di ampiezza.

Si sono quindi calcolati i picchi di massimo per ciascun inviluppo.

Il passo successivo è stato di calcolare la durata di ogni grano: si è scelto – provando diverse ipotesi su grani sintetizzati appositamente – di determinare i tempi d'inizio e di fine, valutando gli istanti all'emivalore. Il calcolo dei campioni che rappresentano gli istanti d'inizio e di fine di un grano è dipeso dalla posizione reciproca assunta dai picchi di massimo relativo, come schematizzato nella figura 19. Per ciascun picco si è proceduto a ricercare il primo campione alla sua sinistra e il primo a destra le cui ampiezze fossero prossime alla metà di quella del picco considerato. Se per i picchi di tipo 'a' i valori sono facilmente trovati scorrendo il vettore contenente i campioni, per quelli di tipo 'b' questo modo di procedere permette di individuare solo il campione d'inizio: per la tipologia 'c' non è possibile determinare l'istante iniziale. Infine, per i grani di tipo 'd' entrambi i valori non possono essere determinati. Si è quindi inserito, in luogo dei campioni mancanti, il valore del picco e si sono associati i grani nella matrice contenente i dati emersi dall'analisi ad un indice associato alle tipologie di figura 19.

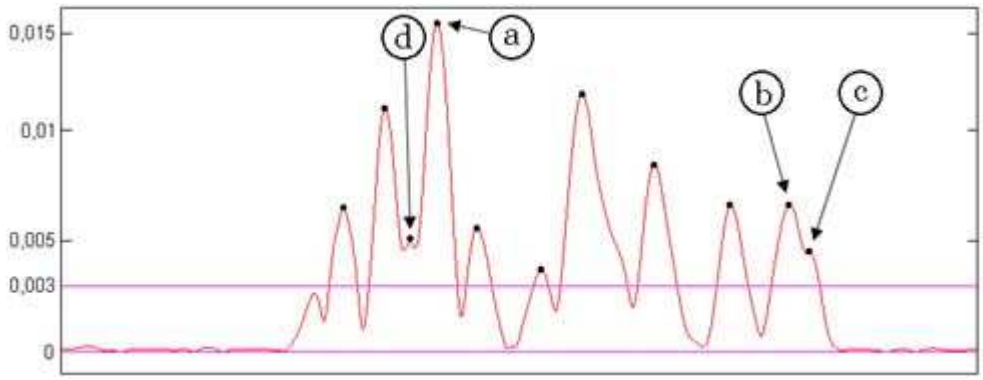



*Fig. 19. Diversa tipologia dei picchi (evidenziati in nero) degli inviluppi dei grani.*

Per ogni frequenza si sono considerati i grani 'a' ed eliminati quelli le cui durate risultano inferiori a 10 ms, valore minimo necessario per percepire il pitch. Per quelli restanti, se ne è calcolata la durata media: si è quindi approssimato con questo valore la durata dei grani di tipo 'b', 'c' e 'd'.

Sono stati quindi definiti 6 diversi livelli di ampiezza, con un intervallo di 6 dB tra loro: questi valori sono stati associati con le notazioni musicali *ff*, *f*, *mf*, *mp*, *p* e *pp*. Poiché i valori di massimo e minimo del segnale sono risultati, rispettivamente, -20 dB e -50 dB, sono stati scelti i livelli -20 dB, -26 dB, -32 dB, -38 dB, -44 dB e -50 dB. Le ampiezze dei grani sono quindi state approssimate al livello più vicino.

La sintesi del brano è stata realizzata utilizzando le frequenze restanti dopo l'analisi (66 disposte nell'intervallo tra 47 e 6331 Hz) con la durata parzialmente corretta, in modo che tutti i grani fossero costituiti da un numero intero di semiperiodi di sinusoide. Quest'accorgimento ha diminuito sensibilmente l'introduzione di spikes percepibili: le variazioni introdotte sono state, nel caso peggiore, di 2.5 ms per grani a bassa frequenza.

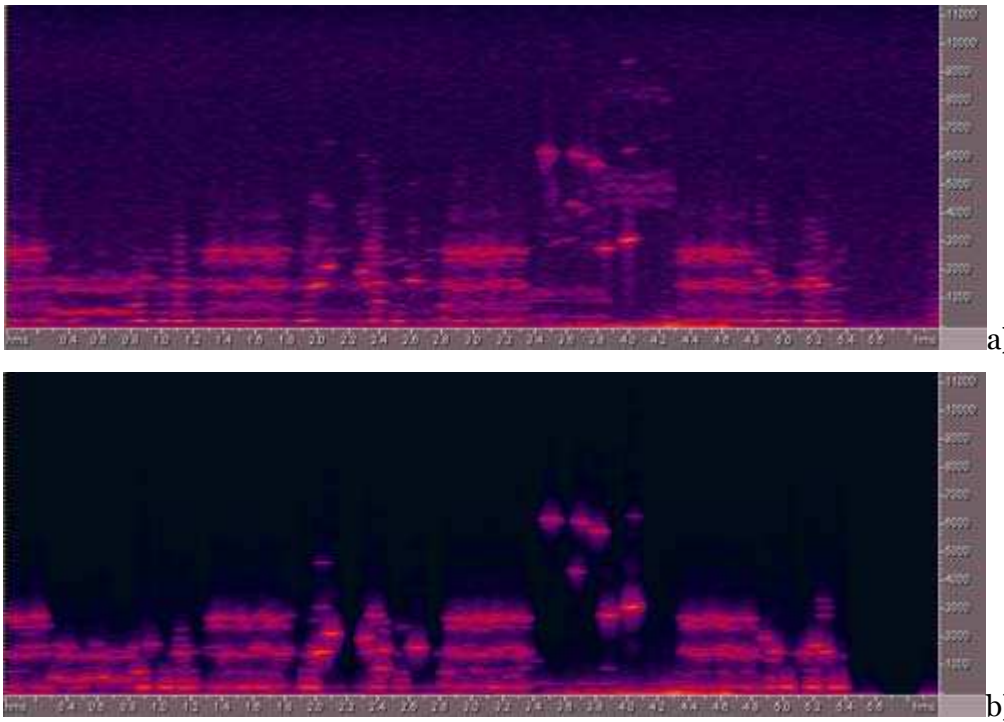
Per ogni grano si è modellata una sinusoide della frequenza e durata desiderate. A questi grani sono stati applicati inviluppi trapezoidali, con rampe di 1 ms – simile quindi all'andamento rettangolare 'teorico', previsto da Xenakis.

La figura 20 mostra i sonogrammi del segnale originale (a) e sintetizzato (b). Il segnale sintetizzato è ritardato rispetto al primo di qualche decina di millisecondi, ma senza alterare la posizione reciproca dei grani.

I 'segnali utili' (i grani, nel caso in esame) delle due immagini sono molto simili, ma si nota, nella figura 18b, l'assenza del rumore. Il segnale ottenuto mediante la sottrazione tra l'originale e il sintetizzato contiene il *rumore* da imputarsi: a) alle imperfezioni e/o all'invecchiamento del supporto; b) alle insufficienze del sistema utilizzato dal compositore. Nell'ipotesi di avere a disposizione gli strumenti musicali (elettronici, nel caso in esame) utilizzati dal compositore, si può pensare di modellizzarli (sintesi a waveguide) e utilizzare questi modelli per la sintesi dell'opera. In questo modo, il suono sintetizzato corrisponderebbe (nel caso ideale) esattamente al suono originale, così come prodotto dal compositore.

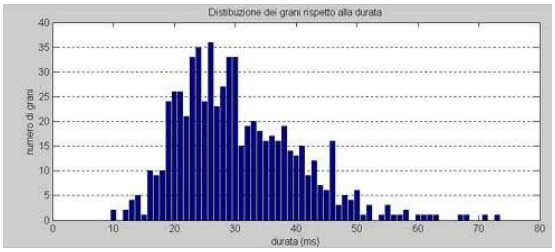
Al fine di verificare le informazioni 'esterne al segnale' sull'opera, si sono effettuate alcune statistiche (fig. 21). La prima (fig. 21a) è relativa alla distribuzione dei soli grani di tipo 'a' (v. fig. 19) rispetto alle durate stimate, il valore medio risulta 30.4 ms. Questa distribuzione smentirebbe l'ipotesi che l'autore abbia utilizzato grani della stessa lunghezza, pari a 40 ms. La seconda statistica (fig. 21b) riguarda la distribuzione

di tutti i grani rispetto alle ampiezze, in dB, stimate durante l'analisi. Si evince chiaramente come non siano presenti solo 4 livelli d'intensità, da cui la scelta, durante la fase di sintesi del segnale, di considerare un maggior numero di livelli.


a)

b)

*Fig. 20: sonogrammi del segnale originale (a) e sintetizzato (b) – finestra di Hanning, lunghezza di 1024 campioni, sovrapposizione del 60%.*

a 

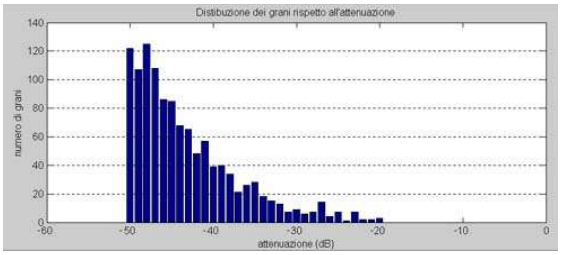

b 


Fig. 21: a) Distribuzione dei grani rispetto alla durata stimata; b) Distribuzione dei grani rispetto all'ampiezza stimata.

Nel segmento esaminato, la densità massima dei grani (massimo numero di grani per secondo), calcolata nell'intervallo del segnale con la maggior concentrazione (tra i 4.31 e 4.85 secondi) risulta essere pari a 370, coerente con uno dei valori indicati da Xenakis.

## Conclusioni

Il problema del restauro di documenti audio è stato affrontato proponendo algoritmi che si basano su approcci diversi tra loro. Sono stati illustrati algoritmi nel dominio della frequenza che rappresentano un'innovazione rispetto ai metodi tradizionali; ad essi sono state aggiunte considerazioni di tipo psicoacustico che permettono di migliorare i risultati ottenuti; infine sono stati presentati metodi nel dominio del tempo che, basandosi su modelli del segnale, consentono di ottenere un notevole miglioramento della qualità del segnale audio.

Per poter confrontare le prestazioni raggiunte dai diversi algoritmi di restauro, si è deciso di introdurre alcuni indici numerici. Va comunque sottolineato che non è possibile ottenere una reale misura numerica in grado di riassumere pienamente il risultato del restauro: in questo senso nulla può sostituire l'ascolto diretto.

Tra le diverse metodologie proposte, va rilevato che i filtri operanti nel dominio del tempo, al prezzo di un'attenta impostazione dei diversi parametri e di un'alta complessità computazionale, consentono di ottenere un restauro in cui il degrado (disturbi impulsivi e rumore a carattere globale) viene attenuato senza inficiare la componente utile del documento audio con un effetto passa basso generalmente presente, al contrario, negli algoritmi che operano nel dominio frequenziale. I filtri nel dominio della frequenza permettono di operare in tempo reale, con un ridotto numero di parametri da regolare. In questo dominio, il filtro EMSR <alpha> permette di ottenere un alto SNR, senza l'introduzione di rumore musicale. In presenza di un SNR minore di 18dB, sorge la necessità di sovrastimare l'impronta di rumore per effettuare una riduzione efficace. Nei brani fortemente degradati, si possono anche studiare opportune combinazioni di filtri in cascata. In questo senso si potrebbe pensare di effettuare una prima elaborazione con un filtro EMSR <alpha> (introducendo un'attenuazione alla maschera di rumore) e una seconda col filtro psicoacustico. In questo modo viene ridotto il rischio di rimuovere componenti utili del segnale.

Non si può comunque pensare di definire una strategia 'ottima' di restauro, in quanto ogni operatore segue i propri principi estetici soggettivi ed è condizionato dall'estetica del periodo storico in cui lavora.

## PER APPROFONDIMENTI


Canazza, S. e Vidolin, A. (2001). Preserving Electroacoustic Music. Special issue of the *Journal of New Music Research*, 30(4).

Canazza, S. (2007). *Noise and Representation Systems: A Comparison among Audio Restoration Algorithms.* Lulu Enterprise, USA. (scaricabile gratuitamente presso: http://www.lulu.com/it/ e acquistabile presso: http://www.amazon.com/)

Ephraim, Y., e Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(6), 1109-1121.

Esquef, P. A. A., Välimäki, V., e Karjalainen, M. (2002). Restoration and enhancement of solo guitar recordings based on sound source modeling. *Journal of the Audio Engineering Society*, 50(4), 227-236.

Godsill, S. J., e Rayner, P. J. W. (1998). *Digital audio restoration – A statistical model-based approach*. London: Springer-Verlag.

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series with engineering applications*. Cambridge, MA: MIT Press.