

Un metodo basato su DSP per l'enhancement dei transienti nella musica registrata

Massimo Magrini*, Tommaso Giunti*, Graziano Bertini*,
Andreas Rauber**, Thomas Lidy**

*ISTI-CNR Pisa, graziano.bertini@isti.cnr.it

** IFS-TU Wien, {rauber,lidy}@ifs.tuwien.ac.at

Sommario

Come è noto le varie tecniche di trattamento del segnale in banda audio nelle radio FM e oggi nella masterizzazione di musica su CD, se da un lato portano ad alcuni vantaggi (rapporto Segnale/Rumore, assenza di clipping, ecc.) dall'altro introducono in molti casi peggioramenti nelle caratteristiche di fedeltà del segnale: ciò a causa delle pesanti compressioni del range dinamico operate, che portano ad una "corposità" e piattezza eccessiva del suono.

E' stato da noi proposto un metodo basato su elaborazione multibanda, mirato a riesaltare i transienti, ampliare la dinamica e la brillantezza del suono in ascolto (Sistema Aria-Digital Dynamic Sound, bibl. Biagiotti-Magrini). Grazie a caratteristiche di flessibilità ed economicità, la procedura opportunamente ingegnerizzata, potrebbe trovare largo impiego come plug-in sw, come modulo hw stand alone, inserito in riproduttori consumer o altri sistemi di retrieval di musica. Nel lavoro si descrive l'hardware del kit di sviluppo DSP adottato per una prima implementazione real-time dell'algorithm.

Un ulteriore funzionalità che è allo studio è quella della possibilità di desumere direttamente dalle features del brano musicale da ascoltare (genere di musica, gradi di compressione) il grado di enhancement ottimale. A questa funzionalità è anche interessato un gruppo dell'Univ. Tecn. di Vienna che si occupa di retrieval da archivi musicali, con il quale è attiva una collaborazione nell'ambito del progetto europeo (MUSCLE-NoE)

A DSP-based method for transient restoration of recorded music

Abstract

The set of audio treatment methods commonly used in many FM radio and in today mastering process of popular music tends to increase the loudness perception of the final audio product, obtaining a loud but "flat" sound. This satisfies the majors for commercial purposes, but the bad effects introduced are dynamic range reduction, poor transients and loss in fidelity at the end. We have proposed a method for time domain transient restoration of recorded music that recovers the clarity similar that of live performance. The algorithm is easily implementable with low cost Digital Signal Processors i.e. in stand alone devices or included in HW (iPod like) or SW (Winamp like) audio player applications. The way to automatically tune the effect by on-the-fly features extraction from the played songs is under study. This goal is pursued jointly with a research group of Tech. Univ. of Wien working in music classification and retrieval, in the framework of EU MUSCLE-NoE project.

1. Introduzione

Il sistema ARIA – Digital Dynamic Sound propone un’innovativa strategia, per l’elaborazione dei segnali audio digitali, finalizzata all’arricchimento della dinamica che l’ascoltatore percepisce durante la riproduzione di un CD musicale o altra sorgente digitale. L’obiettivo è quello di cercare di ripristinare le caratteristiche originarie in termini di dinamicità, spazialità e completezza armonica che spesso vengono modificate (e a seconda del tipo di musica ooggiorno notevolmente deteriorate) in fase di missaggio e masterizzazione.

Infatti l’artificio della compressione della dinamica nella musica registrata (su CD e altri supporti), usato molto correttamente in passato, negli ultimi anni è stato abusato (secondo opinione corrente) a causa di pure ragioni commerciali.

Da diversi anni nella radiofonia in modulazione di frequenza, specie nelle emittenti private, vengono usati dei “pesanti” compressori audio prima di inviare i segnali ai trasmettitori, in modo da invogliare l’ascoltatore in ricezione a “fermarsi” su stazioni che suonano più “forte e chiaro” (ad es. per sovrastare eventuali rumore d’ambiente). Seguendo l’idea che più ascoltatori significa più introiti pubblicitari, in poco tempo l’effetto è stato portato al suo limite massimo in quasi tutte le stazioni, tanto che le persone si sono abituate ad ascoltare la musica in questo modo. Inevitabile quindi il successivo trasferimento di questa tendenza direttamente al supporto registrato, che altrimenti “suonerebbe” molto meno corposo rispetto al segnale radio, con timori di ricadute negative sul mercato discografico. In questo modo la dinamicità dei transienti che rende vive le registrazioni ben fatte, viene a mancare in gran parte delle produzioni discografiche moderne (vedi [1L] web-site e [9])

Nel 1999 il DJ Gabriele Biagiotti, a cui si può attribuire la paternità di ARIA, si accorse che il suono si “impreziosisce” quando si “insegue”, esaltandolo con i cursori dell’equalizzatore, l’andamento dell’ampiezza di ciascuna banda audio, osservato su un analizzatore di spettro in tempo reale. In maniera empirica, Biagiotti ricostruiva in qualche modo i transienti mancanti nelle registrazioni. Riscontrando un interesse positivo negli ascoltatori per questa manipolazione, immaginò che un dispositivo (HW/SW) potesse eseguire quell’effetto in modo più efficace e, cosa più importante, automaticamente.

Lo sviluppo tecnico del metodo è stato affrontato presso il laboratorio Segnali&Immagine dell’I.S.T.I. del C.N.R. di Pisa, con uno studio di fattibilità e con la realizzazione della prima versione sperimentale del sistema. Massimo Magrini trovò che l’effetto di enhancement sulla dinamica si poteva ottenere rilevando ed elaborando opportunamente le variazioni di ampiezza dei segnali nelle varie sottobande audio; l’equalizzazione del segnale sarebbe stata fatta dipendere dalle variazioni rilevate. Sulla base di queste prime considerazioni, è stato sviluppato un algoritmo e implementato su un prototipo del sistema utilizzando la tecnologia DSP, con deposito di brevetto (n. 3957PTIT/2, 2003 [1]).

Successivi approfondimenti con lo scopo di testare altre soluzioni per l’algoritmo e di ottimizzarne alcuni aspetti, (tempo di esecuzione, maggiore generalità d’uso ecc.) sono stati effettuati nell’ambito di una tesi di laurea, come pure è in sviluppo una versione preliminare dell’algoritmo in forma di plug-in per Winamp. Alcuni risultati sono stati presentati in meeting scientifici nell’ambito del progetto europeo (MUSCLE-NOE Multimedia Understanding through Semantics, Computation and Learning, FP6 – 507752), attivo su temi di ricerca relativi a classificazione, archiviazione e retrieval di eventi audio/video.

Un’ulteriore funzionalità che è allo studio è quella della possibilità di desumere direttamente dalle features del brano musicale da ascoltare (genere di musica, livelli di compressione) il grado di enhancement ottimale, che nella versione attuale viene regolato empiricamente. A questa funzionalità è anche interessato un gruppo dell’Università IFS-TU Wien, che si occupa di classificazione e retrieval da archivi musicali, con il quale è attiva una collaborazione sempre nell’ambito del progetto MUSCLE-NOE.

2. Richiamo alle operazioni sui brani musicali eseguite in fase di masterizzazione

Come è noto, a partire dalle vecchie tecnologie analogiche, sono stati applicati vari metodi di ottimizzazione nelle registrazioni di segnali musicali nei riguardi del rapporto segnale rumore, clipping ecc.: basti pensare ai diffusissimi sistemi Dolby, con tutte le varianti ed evoluzioni, per le quali esiste una notevole letteratura tecnica. Attualmente per le operazioni di preprocessing e masterizzazione vengono utilizzate nella stragrande maggioranza dei sistemi commerciali e professionali le tecniche numeriche: il “segnale” viene condizionato, filtrato, compresso in modo tale da soddisfare una serie di esigenze ad es. renderne la dinamica compatibile con quella del supporto su cui verrà registrato ecc. Qui di seguito riportiamo un breve richiamo ai principali tipi di elaborazione che sono

applicate, dando solo qualche dettaglio sulla *compressione* e rimandando alla bibliografia per una completa panoramica sulle tecniche del mastering audio [2].

Normalizzazione - Tale processo incrementa l'ampiezza complessiva del segnale fin quando il campione più grande non raggiunge il livello di clipping, facendo coincidere la dinamica di quantizzazione con quella del segnale.

Equalizzazione - Attraverso il processo di equalizzazione è possibile bilanciare il contributo armonico del segnale: tendenzialmente oggi si cerca di ottenere uno spettro del segnale piatto, con un eventuale leggero boost sulle basse frequenze.

Compressione - Durante i passaggi musicali molto dolci (piano o pianissimo, in termine tecnico) una buona parte della definizione sonora verrebbe irrimediabilmente persa se si usa la stessa risoluzione che in tutti gli altri intervalli di dinamica del segnale. Tenendo conto di questo e di altri fattori fra cui il fatto che oggi la musica viene ascoltata in ambienti di regola non silenziosi, in fase di pre-masterizzazione viene applicato un meccanismo che tende ad aumentare l'ampiezza del segnale quando esso è di livello basso, ed invece a diminuirla se essa supera una certa soglia.

Questo artificio è riferito come *compressione della dinamica* (vedi Fig. 1) e viene applicato sia sulle singole tracce durante la lavorazione del brano, sia sul mix finale: in pratica si riducono quindi i livelli più alti della forma d'onda del segnale mentre si esaltano quelli più piccoli.

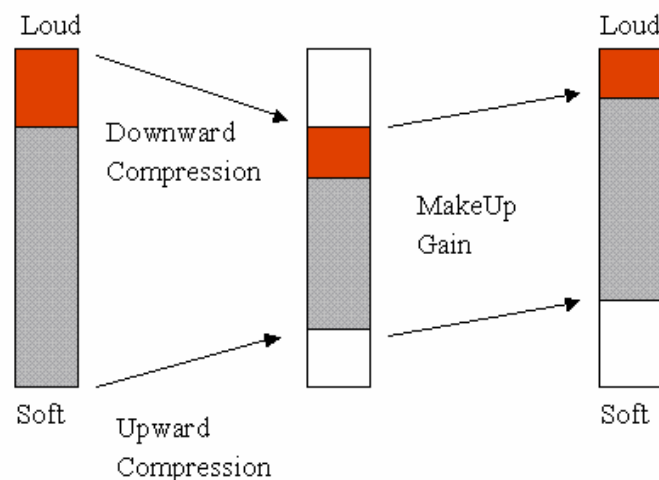


Figura 1. Compressione tipica del range dinamico durante il mastering

Massimizzazione - La maggior parte dei segnali presenta picchi in ampiezza: in corrispondenza di particolari passaggi con il cantato più intenso o attacchi molto incisivi. È stato escogitato un artificio il cosiddetto "Look Ahead Peak Limiting" che riduce solamente i picchi di volume più alti senza compromettere il resto del segnale. Questo avanzato algoritmo di massimizzazione del segnale, abbinato ad una sofisticata tecnologia di riquantizzazione, permette un notevole aumento di quello livello efficace medio (RMS), ottenendo un segnale più robusto e uniforme.

3. Effetti della compressione nel riascolto.

La tendenza attuale, per vari motivi già citati, è quella ad esasperare l'effetto della compressione e della massimizzazione, in una competizione verso il volume più "alto" possibile, che spesso però va a scapito della qualità e della fedeltà.

Sebbene una moderata compressione abbia effetti positivi, una compressione pesante, oltre i 12 dB, rende la musica innaturale per chi è abituato alle dinamiche delle esecuzioni *dal vivo* o di buone registrazioni di musica classica. Il suono risulta "appiattito", monotono, impoverito in termini di fedeltà e spazialità: il valore RMS di un

brano rock con 30 dB di dinamica originale, dopo essere stato masterizzato sui comuni CD, risulta contenuta generalmente entro il valore di -6 dB (vedi Fig 2).

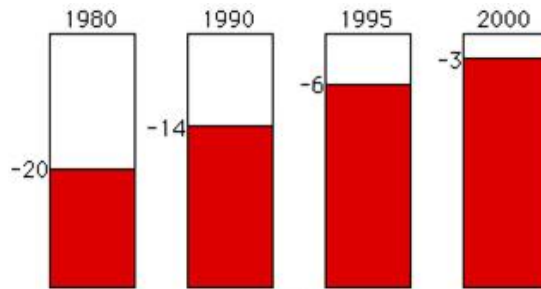


Figura 2. Cambiamento nel volume percepito nella musica commerciale dal 1980 a oggi.

Nella musica, la dinamica è una peculiarità di grande valore: ne accentua l'emozionalità creata dal contenuto. Pensiamo alla colonna sonora di un film, che sui DVD viene digitalizzata e compressa come sopra illustrato. Picchi di intensità creano suspense e sottolineano i colpi di scena; tutto ciò viene drasticamente ridimensionato nell'uso esasperato che si fa oggi con il processo di compressione e massimizzazione.

L'ascoltatore attento alla qualità sente dunque l'esigenza di eliminare il senso di piattezza e quindi sembra giustificato il bisogno di attuare metodi in grado di restituire alla musica quelle caratteristiche di cui è stata privata.

4. L'algoritmo ARIA

L'algoritmo di elaborazione ARIA – Digital Dynamic Sound permette di esaltare le transizioni di volume del suono e cercare di ripristinare le caratteristiche originarie in termini di dinamica, spazialità e anche completezza armonica.

Concettualmente si usa un criterio opposto a quello della compressione, però applicandolo in modo dinamico e non statico come nei classici espansori.

Il segnale viene dapprima processato con un banco di filtri e scomposto in dieci sotto-bande, ciascuna elaborata in modo indipendente prima di essere nuovamente miscelata (vedi Fig. 3).

Il procedimento consiste nel percepire le variazioni d'intensità del segnale nelle singole bande mediante il confronto tra due tipi di inviluppi, ricavati opportunamente a run-time (vedi oltre) con due velocità di risposta differenti. Dal confronto dei due inviluppi si ricava un indice di "variabilità istantanea" che viene usato come fattore di amplificazione del segnale stesso.

Diagramma a blocchi – L'algoritmo può essere adattato con estrema flessibilità ad ogni moderno sistema di riproduzione audio: mono, stereo o multicanale, a prescindere dalle sue caratteristiche. Per semplicità, prenderemo in considerazione un solo canale ed un'unica banda.

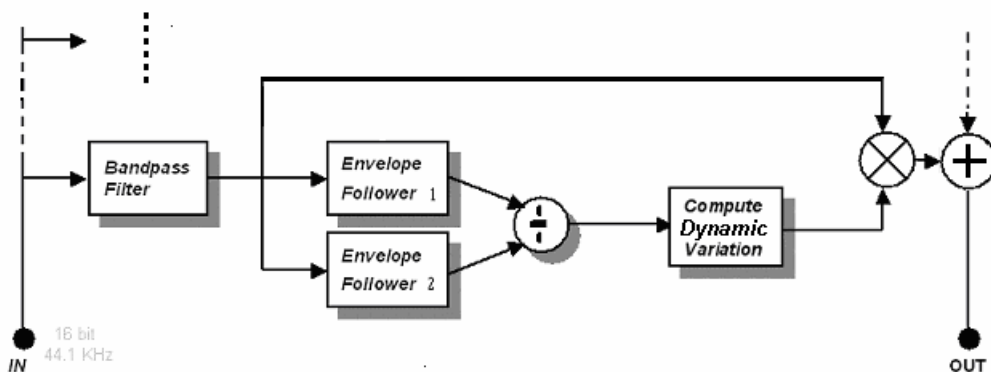


Figura 3. Schema a blocchi dell'algoritmo

In riferimento alla figura, i campioni del segnale in ingresso vengono innanzitutto filtrati da un banco di filtri digitali IIR (Infinite Impulse Response).

Sperimentalmente, nel nostro prototipo si è notato che il numero di filtri (sottobande) ottimale, in termini di qualità e complessità computazionale è dieci (vedi tabella 1), caratterizzati dal fatto che ciascuno ha frequenza di centro banda doppia rispetto al filtro precedente.

La risposta in frequenza di un filtro passa-banda selettivo ideale può essere approssimata ad es. disponendo in cascata due filtri Butterworth del secondo ordine di cui uno passa-basso e l'altro passa-alto con opportune frequenze di taglio (vedere a par. 7 un esempio di implementazione su DSP basato su un'altra tecnica).

Tab. 1

$f_c 1$	$f_c 2$	$f_c 3$	$f_c 4$	$f_c 5$	$f_c 6$	$f_c 7$	$f_c 8$	$f_c 9$	$f_c 10$
31,5Hz	63Hz	125Hz	250Hz	500Hz	1KHz	2KHz	4KHz	8KHz	16KHz

Del segnale in uscita dal filtro della banda i -esima a noi interessa seguirne l'andamento del volume o meglio la velocità con cui cambia l'ampiezza e per questo ci serviranno i due tipi di involuপি. Per questa operazione si possono usare diversi metodi [3, 4].

a) Ad es. si può elevare ciascun campione al quadrato e filtrare poi il risultato con un semplice filtro numerico passa-basso del primo ordine. Scegliendo la larghezza di banda si ricavano degli involuপি (uno a tempo breve e uno a tempo medio) che "seguono" con diverso grado di smoothing le variazioni istantanee del segnale. Un filtro con banda stretta rende il rivelatore di involuপি meno "reattivo", cioè rallenta la velocità di risposta alle variazioni. La scelta della frequenza di taglio non è critica, ed è determinata tenendo conto del compromesso tra velocità di risposta e caratteristiche di smoothness dell'involuপি rivelato.

b) Un'altra strategia per seguire l'andamento dell'involuপি di un segnale consiste nell'integrare il quadrato dei campioni all'interno di una finestra temporale mobile che nel dominio digitale si riduce nel calcolo di una media operata su un certo numero di campioni M , ed esprimibile come:

$$acc_i[n] = \frac{1}{M} \cdot \sum_{k=0}^{M-1} c_i^2[n-k] \quad (1)$$

Rispetto alla precedente, questa strategia si dimostra più efficiente in termini di smoothness, soprattutto sulle bande più basse, le cui frequenze sono dell'ordine della frequenza di taglio del corrispondente filtro passa-basso dell'altro metodo (vedi fig. 4).

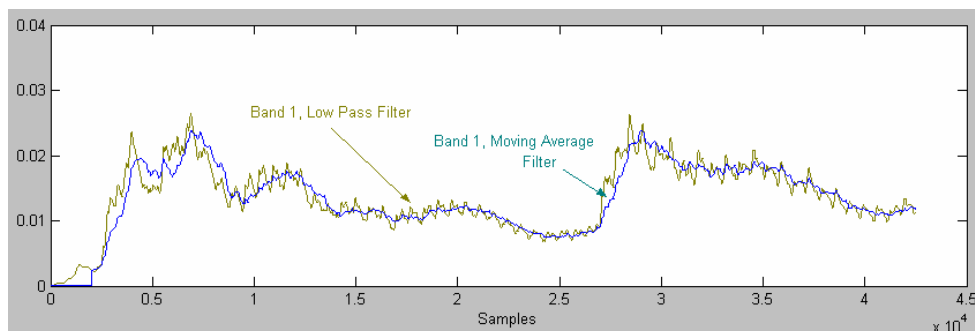


Figura 4 . Confronto fra gli involuপি ricavati con i due metodi per una banda

Il filtraggio a media mobile consiste nella convoluzione del segnale con un impulso rettangolare discreto di ampiezza $1/M$: la cui risposta in frequenza è:

$$H(f) = \frac{\sin(\pi fMT)}{M \cdot \sin(\pi fT)} \quad (2)$$

e la cui banda a -3dB equivale a $B_{-3} \cong 0,3/(M \cdot T)$. Si osserva quindi che il parametro M influenza in modo inversamente proporzionale la velocità di risposta del rivelatore di inviluppo. Infatti al crescere di M , la finestra temporale si allarga e il numero di campioni da mediare aumenta: l'andamento dell'inviluppo subirà scostamenti ridotti attorno ad un livello che rappresenta il valor medio del segnale. Viceversa ridurre il valore di M significa analizzare una finestra più stretta, il che contribuisce a dare più peso alle transizioni (vedi Fig. 5 con $M_2 > M_1$).

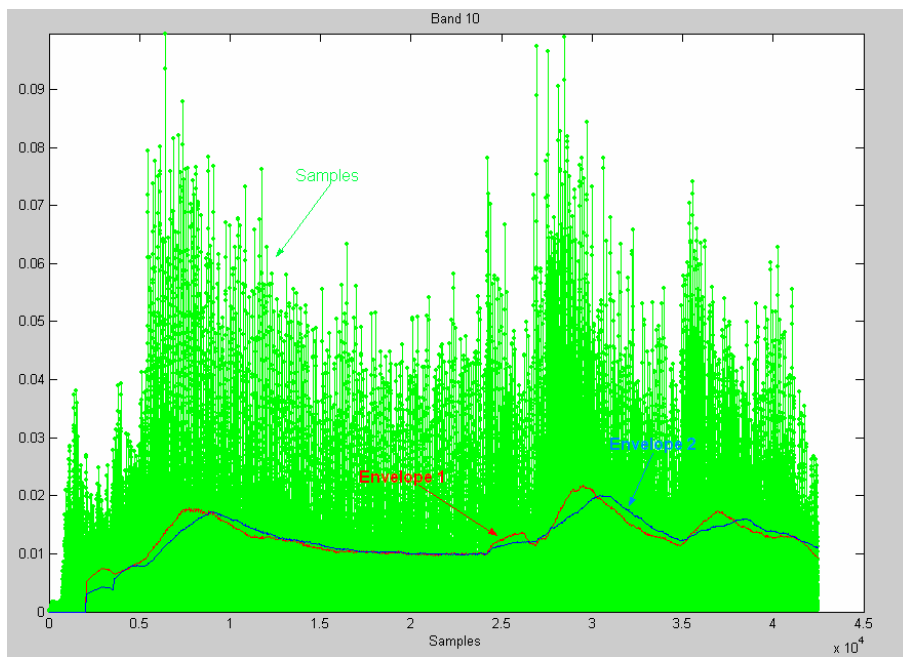


Figura 5 - Inviluppo a tempo breve (1) e medio (2) con filtro a media mobile

Dalla figura si osserva chiaramente che in corrispondenza di un transiente d'attacco, dove il volume è crescente, l'inviluppo con velocità di risposta maggiore (poniamo $acc[n]$) incrementa il proprio valore più di quanto non faccia l'inviluppo lento (poniamo $acc^{long}[n]$) e viceversa nella fase di rilascio. Inizializzando $acc[n]$ e $acc^{long}[n]$ con lo stesso valore e calcolando il rapporto:

$$R_i[n] = \frac{acc_i[n]}{acc_i^{long}[n]} \quad (3)$$

per ogni n (cioè ad ogni periodo di campionamento) si ottiene un indice della variabilità del volume della singola banda i -esima. Se l'indice $R_i[n]$ è maggiore di uno, il livello istantaneo del segnale si sta incrementando, se è minore di uno, sta diminuendo, mentre se è prossimo ad uno è da ritenersi pressoché stazionario.

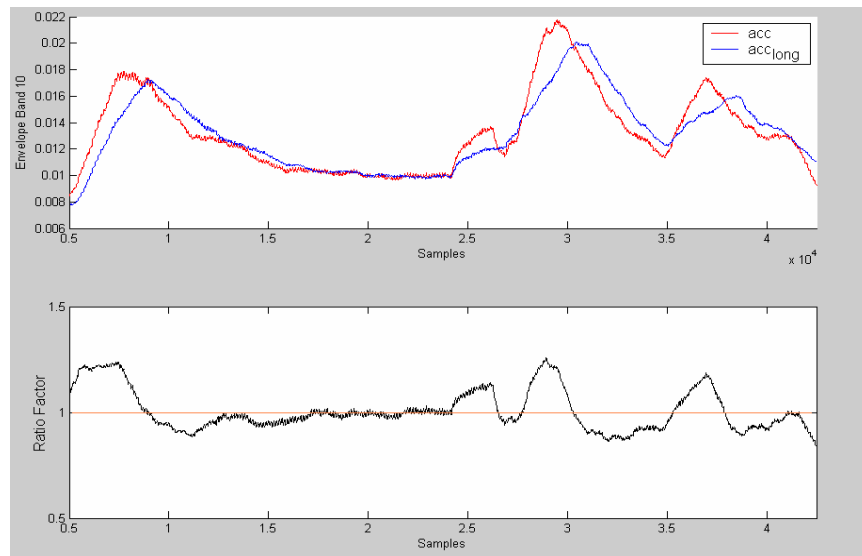


Figura 6 Esempio dell'andamento di $R_i[n]$

L'indice di variabilità del volume $R_i[n]$ viene usato come moltiplicatore per il campione che si trova all'uscita del filtro della banda i -esima:

$$out_i[n] = c_i[n] \cdot R_i[n] \quad (4)$$

Se l'involuppo è in prossimità di un fronte in salita, i corrispondenti campioni vengono amplificati di un fattore proporzionale alla rapidità dell'incremento. Se l'involuppo è in prossimità di un fronte in discesa, i corrispondenti campioni vengono proporzionalmente ridotti. Si ottiene così un segnale con le variazioni dinamiche istantanee di volume opportunamente accentuate (vedi Fig. 7).

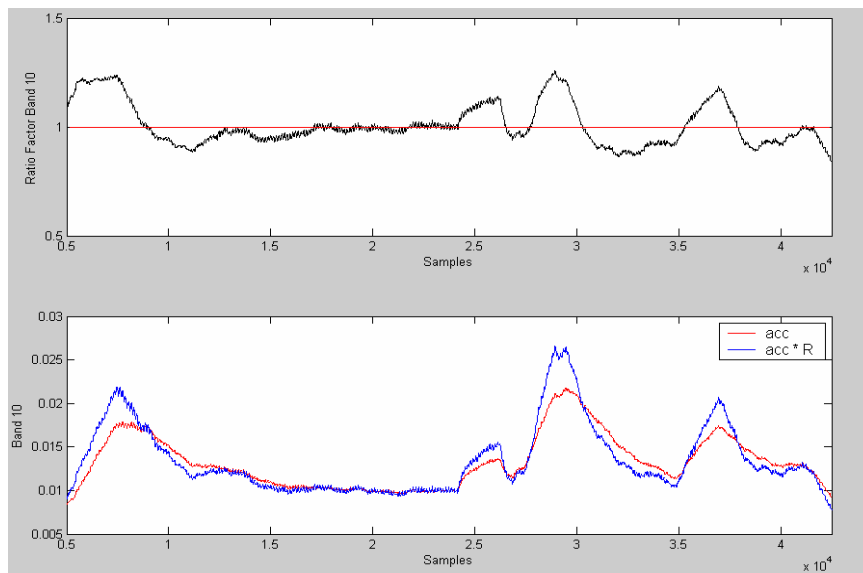


Figura 7. Involuppo del segnale in uscita elaborato con $R_i[n]$

L'indice di variabilità del volume $R_i[n]$ può assumere qualunque valore maggiore di zero, anche molto discosto dall'unità. Nel caso di segnali molto ritmici, con veloci variazioni dinamiche, si raggiunge anche l'ordine della decina: applicando un'amplificazione così elevata (è stato riscontrato sperimentalmente) rende il risultato innaturale e va ben oltre l'obiettivo di compensare le elaborazioni descritte nel capitolo precedente. Per cui si

rende necessario limitare tale indice nei suoi valori massimi e minimi. Dei valori tipici sono 2 e 0.5: se il rapporto è superiore o inferiore a queste soglie viene opportunamente limitato a tali valori.

Così facendo si genera un effetto che ha, al massimo, il peso di ± 3 dB sul segnale originale. Tuttavia la scelta dei limiti non è critica e può essere fatta a discrezione dell'utente in base al genere di musica, ecc.

Il processo appena descritto viene eseguito separatamente sui singoli canali e indipendentemente su ciascuna delle dieci bande dello spettro audio: il contributo sul campione d'uscita all'istante presente sarà dato dalla somma:

$$OUT [n] = \sum_{i=0}^{10} out_i [n] \quad (5)$$

Alcuni esempi dell'effetto sulle ampiezze d'uscita su un paio di sottobande sono riportati in Appendice.

Differenza con un Espansore classico – Al contrario di un normale espansore di dinamica, ARIA agisce soltanto in corrispondenza delle transizioni dinamiche delle singole bande e non sui valori assoluti. Due campioni della stessa ampiezza vengono trattati diversamente a seconda della regione dell'involuppo a cui appartengono. Se l'involuppo presenta un andamento costante, il campione non viene modificato; qualora ci si trovi in presenza di un transiente d'attacco o di rilascio, il campione sarà alterato in maniera da enfatizzarne l'effetto.

Il comune espansore invece utilizza una tabella di conversione: la profondità del suo intervento è legata in modo continuativo nel tempo solo al livello del segnale (vedi Fig. 8). Al contrario, il processore ARIA interagisce sulle variazioni istantanee senza alterare significativamente né il contenuto timbrico totale, né il volume medio.

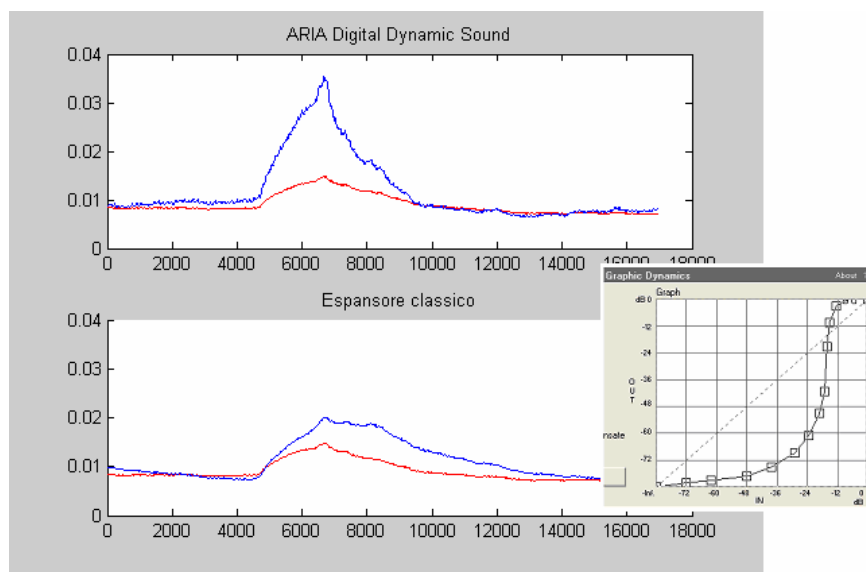


fig. 8. Differenza tra un espansore standard e l'algoritmo ARIA

6 Implementazione su DSP

Una versione prototipale del sistema ARIA è stata realizzata su piattaforma Hardware digitale grazie ad un Digital Signal Processor con specifico firmware. E' stata impiegata la scheda DSK 21061 Ez-KIT Lite prodotta da Analog Device. Nell'ottica di eventuali destinazioni commerciali del sistema, si è cercato di realizzare un prototipo economico con rigidi vincoli nello sviluppo del sistema, in termini di risorse computazionali e di memoria. E' stato necessario adeguare l'algoritmo affinché l'elaborazione si completasse all'interno di un periodo di campionamento (real-time).

E' stato introdotto un sottocampionamento prima dei rivelatori di involuppo, tale da non compromettere il verificarsi della condizione di Nyquist. E' importante sottolineare come l'enhancement della dinamica coinvolga comunque l'intera sequenza a 44.1 KHz; il sottocampionamento aggiorna, con periodi diversi banda per banda, l'indice di variabilità e ciò è congruente con la velocità di variazione della relativa banda.

Il calcolo del rapporto, particolarmente pesante in termini di risorse di calcolo, è stata semplificato con una moltiplicazione:

$$R_i[n] = \frac{acc_i[n]}{acc_i^{long}[n]} = acc_i[n] \cdot \left(\frac{1}{acc_i^{long}[n]} \right) \quad (6)$$

Normalizzando tra 0 ed 1 il valore dei campioni in ingresso, il termine $1/(acc_i^{long}[n])$ può essere ottenuto per tabellizzazione della funzione $1/x$. Così facendo si ottiene una buona approssimazione del calcolo richiesto con l'accesso ad un vettore e ad una moltiplicazione, evitando la divisione, operazione pesante per i DSP, per la quale non sono ottimizzati.

Un altro passaggio critico dell'algoritmo è il filtraggio del segnale in sottobande: l'uso di due filtri IIR Butterworth (LP e HP in cascata) del secondo ordine risulta troppo oneroso. Si è scelto dunque un compromesso tra velocità computazionale e selettività implementando i filtri con una funzione biquadratica con andamento a "campana".

Nel caso di sistemi multicanale viene calcolato un solo indice sulla variabilità media dei canali. Ciò è plausibile se si considera lo stretto vincolo di correlazione che sussiste tra le ampiezze dei due canali. Tuttavia, è importante evidenziare come le peculiarità di stereofonia del segnale non vengano minimamente compromesse.

La scarsa disponibilità di memoria nel DSP adottato rende necessario rivelare l'involuppo a tempo più lungo con il filtro LP. Infatti il filtro a media mobile necessiterebbe di un numero elevato di campioni e se volessimo calibrare l'effetto correggendo il calcolo dell'involuppo lento dovremmo garantire lo spazio in memoria per il caso peggiore, ciò comporterebbe uno spreco di risorse in tutti gli altri casi. Invece con il filtro LP è necessario allocare in memoria solamente un valore per ciascuna banda: quello dell'involuppo all'istante precedente; per di più, la correzione del calcolo dell'involuppo lento viene fatta agendo soltanto sul valore delle costanti e non sull'allocazione di memoria.

Impieghi e sperimentazioni - Il sistema ARIA – Digital Dynamic Sound, con un'adeguata ingegnerizzazione, potrebbe diventare un prodotto consumer che, per le sue caratteristiche e semplicità d'uso, potrebbe essere integrato nei comuni apparati di riproduzione musicale.

Un prototipo del sistema è stato realizzato in versione hardware basata su tecnologia DSP: la scheda è stata installata all'interno di un case (1 unità rack 19") con ingressi-uscita RCA. È stata dotata di alimentazione indipendente (versione stand-alone) che può essere collocata nella catena di un sistema Hi-Fi permettendo un incremento della qualità dell'ascolto (vedi Fig. 9).

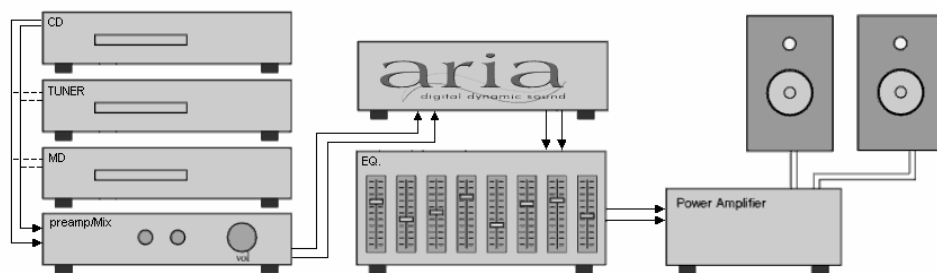


Figura 9. Disposizione di un rack ARIA in un sistema di riproduzione

Con un'adeguata ingegnerizzazione le funzioni del sistema sarebbero integrate al limite con un unico chip di piccole dimensioni, facilmente inseribile in amplificatori, speaker preamplificati, autoradio, lettori di CD-ROM, MP3 player, sistemi Home-Theatre, etc..

Modulo plug-in - E in perfezionamento una realizzazione software dell'algoritmo (con enhancement prefissato o regolabile dall'utente) per l'elaborazione real-time di file audio memorizzati nelle memorie di massa del personal computer costituito da un modulo plug-in DSP per software quali Winamp o simili (attualmente funzionante in versione preliminare).

7. Sviluppi

Adattamento del fattore di guadagno in modalità automatica. Una direzione di ricerca è attualmente in corso di svolgimento con l'obiettivo di definire in modo automatico il fattore di enhancement in base alle caratteristiche specifiche del brano in ascolto. Quello che servirebbe è una stima del grado di compressione di

un brano in ingresso, ottenendola con analisi degli andamenti delle ampiezze dei segnali, sempre in opportune sottobande.

Come prima soluzione, sembra possibile stabilire l'entità dell'enhancement operando una trasformazione lineare dell'indice $R_i[n]$ definito in precedenza nella (6). Ciò richiede una preanalisi a run-time del segnale musicale per stimare le proprietà statistiche della variabile aleatoria $R_i[n]$, le escursioni del contenuto energetico e, come vedremo, poter sfruttare la conoscenza del genere musicale del brano in riproduzione.

La varianza di $R_i[n]$ fornisce in effetti una misura della variabilità del segnale: valori prossimi a zero caratterizzano distribuzioni concentrate attorno a $R_i[n]=1$ cioè segnali lentamente variabili; invece segnali tanto più briosi sono caratterizzati da una varianza tanto più maggiore di zero. Tale varianza può essere stimata, ed aggiornata, con sufficiente precisione in un numero contenuto di osservazioni (campioni). ($0,1 \text{ sec}=4410 \text{ samples} \rightarrow \text{Root Mean Square Error} < 10E-3$). La trasformazione lineare su $R_i[n]$ ha lo scopo di aumentarne la varianza attraverso step progressivi fino al raggiungimento di un valore prefissato: questo permette di avere sempre lo stesso guadagno in dinamica a prescindere dalla variabilità del segnale in ingresso. L'algoritmo

determina un valore di soglia r_i^* , massimo guadagno, scegliendolo come compromesso tra la probabilità di clipping tollerabile e $f_{R_i[n]}(r_i^*)$ (Probability Density Function) non troppo piccolo. In pratica è necessario che

vengano prima attenuate tutte le sottobande di un fattore costante pari a $[1/r_i^*]_{max}$ (per sicurezza) per poi essere trattate singolarmente secondo il guadagno tempo-variante $R_i[n]$. Così facendo si ottiene un ampliamento della dinamica, intesa come differenziale tra i valori di picco possibili e il valore RMS finale, approssimativamente dell'ordine di $20 \cdot \log_{10} r_i^*$ dB (vedi Fig 8).

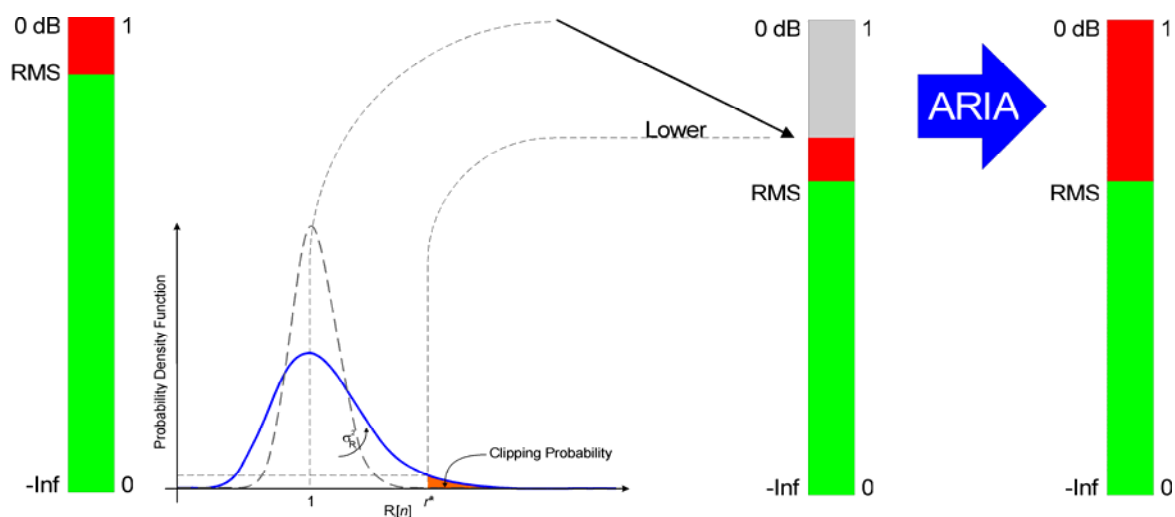


Figura 8. Situazione dei livelli di ampiezza prima e dopo l'aggiustamento.

Adattamento del fattore di guadagno mediante analisi semantica - Un'altra direzione di ricerca, attualmente in corso di studio, si basa sui moderni paradigmi della semantica audio e della classificazione per generi musicali come supporto per l'applicazione automatica dell'enhancement. La tipologia di operazioni, descritte nel seguito, si inquadra tra l'altro nell'attività svolta nell'ambito della rete di istituti del citato progetto MUSCLE-NOE in cui gli autori sono coinvolti. I temi affrontati riguardano le problematiche tipiche della classificazione e del retrieval di dati da archivi audio e musicali sulle quali è in particolare attiva l'unità di ricerca presso l'Università di Vienna (IFS-TU Wien) [5] nostro partner sullo specifico tema "Semantic from audio" (sito web [7L]).

Il gruppo IFS-TU ha sviluppato un sistema denominato PlaySOM [5L, 6] che rende conveniente l'accesso ad un archivio musicale mediante un speciale interfaccia grafica che mappa su una superficie bidimensionale i vari brani, raggruppandoli in zone per similarità di caratteristiche (features) e quindi di genere. La costruzione dell'archivio è ottenuta off-line con una serie di operazioni e di passaggi piuttosto complessi e articolati (descritti in dettaglio in [5,7] di cui qui citiamo i principali:

- 1) analisi ed estrazione di un opportuno numero di features
- 2) classificazione dei generi tramite l'ambiente WEKA -Machine Learning Software [6]
- 3) creazione dell'interfaccia utente tramite mappe bidimensionali applicando il criterio delle Self-Organizing Map-SOM [8].

Il passo numero 1 è quello che ci interessa da vicino e fa uso di varie operazioni di Digital Signal Processing sui segnali di un brano musicale codificato in vari formati: vengono infatti estratte una serie di features tramite opportune STFT e FT che analizzano la distribuzione temporale dell'energia in 24 sottobande (bande critiche). Tra le principali features che sono considerate ci sono quelle del *Rythm Patterns*, *Rythm Hystogram and Statistical Spectrum Descriptor*.

Un'ulteriore feature potrebbe essere la varianza di $Ri[n]$, come da noi definita in precedenza, nelle varie sottobande che, dopo durante la classificazione, verrebbe ad essere stimata come "tipica" di un genere. La riproduzione di un brano dell'archivio di PlaySOM potrebbe quindi, se desiderato dall'utente, attivare e far eseguire un modulo che implementa ARIA come descritto nel punto precedente, però senza bisogno di fare un'analisi del segnale on-fly.

Questa strategia avrebbe margini di miglioramento e assumerebbe maggiore validità e rilevanza generale conoscendo i livelli di $Ri[n]$ e anche degli RMS statistici di riferimento per ciascun genere, ma questo non può essere determinato analizzando un ristretto archivio di musica registrata qualsiasi, bensì un più vasto archivio con brani di vecchie registrazioni non manipolate o anche attuali, ma prima della fase di mastering, o meglio ancora registrati dal vivo. Un tale studio consiste nell'analizzare, per una durata sufficiente, decine o centinaia di brani musicali di generi diversi e per ottenere dati statisticamente rilevanti sarebbe necessario un'attività troppo lunga e sistematica che esula dalle nostre attuali possibilità [9].

8. Conclusioni

L'avvento della tecnologia digitale ha senza dubbio portato enormi benefici nel campo musicale, sia sotto l'aspetto della fruibilità, che dal punto di vista dell'elaborazione del segnale. Tuttavia, con il tempo, è cambiato anche il modo di pensare ed apprezzare la qualità del suono; stiamo assistendo ad un fenomeno generalizzato in cui "l'alta fedeltà" non è più perseguita come obiettivo primario e ciò viene anche sottolineato con preoccupazione da molti esperti (vedi "The death of dynamic range" [1L]).

Il sistema ARIA – Digital Dynamic Sound si inserisce in questo contesto contribuendo a reintrodurre un certo amount di contenuto dinamico della musica (agendo sulla dinamica dei transienti) che sempre più spesso viene pesantemente artefatto in fase di masterizzazione dei media digitali. L'elaborazione è multibanda e real-time, in grado di processare anche sorgenti multicanale. L'algoritmo è flessibile in quanto offre la possibilità di personalizzare l'incisività dell'elaborazione semplicemente agendo su un solo parametro. È attualmente allo studio la modalità di applicazione automatica dell'effetto, basandosi sui moderni paradigmi della semantica audio e della classificazione per generi musicali ecc. ed altri eventuali miglioramenti, che potrebbero essere implementati in real-time qualora si disponga di una piattaforma DSP sufficientemente potente.

Ringraziamenti

Si ringrazia Gabriele Biagiotti per avere suggerito il tema del presente lavoro e partecipato ai test audio dell'algoritmo. Il lavoro si avvale anche del supporto del progetto EU MUSCLE NoE -Multimedia Understanding through Semantics, Computation and Learning- contract 507752 nell'ambito del 6° FP, IST.

Bibliografia

- [1] Biagiotti G., Magrini M., *Documentazione del brevetto Aria release n. 3957PTIT/2 -2003*.
- [2] Katz B., *Mastering Audio, The Art and the Science*. Focal Press Book, 2002.
- [3] Luise M., Vitetta G., *Teoria dei segnali*. McGraw-Hill, 1999.
- [4] Tarabella L., Bertini G., Caioli A., Guerra A., *Informatica e Musica*. Jackson Libri, Milano 1992.
- [5] Lidy T., Rauber A., "Evaluation of Feature Extractors and Psycho-acoustic transformations for Music Genre Classification" in Proc. of the Sixth Int. Conf. on Music Information Retrieval (ISMIR 2005), pp. 34-41, London, UK, Sep. 11-15, 2005.
- [6] Neumayer R., Dittenbach M., Rauber A.. *PlaySOM and PocketSOMPlayer: Alternative Interfaces to Large Music Collections*. Proc. of the Sixth Int Conf on Music Information Retrieval (ISMIR 2005), pp. 618-623, London, UK, September 11-15, 2005.
- [7] Neumayer R., Lidy T., Rauber A. *Content-based Organization of Digital Audio Collections* Proc. of the 5th Open Workshop of MUSICNETWORK, Vienna, 2005.
- [8] Kohonen T. *Self-Organizing Maps*. Springer, 2003.
- [9] Noselli G. *Acustica, Psicoacustica e Dintorni* web site-CD Mastering Services.

Links utili

- [1L] www.cdmasteringservices.com *The Death of Dynamic Range* (no author); Speer B. *What happened to Dynamic Range?*
- [2L] www.analog.com Analog Device Inc *SHARC Ez-KIT Lite Document Library 1997, ADSP 21061 Data Sheet*. 1996
- [3L] www.musicdsp.org
- [4L] www.aria99.com

[5L] <http://www.ifs.tuwien.ac.at/~andi/mir/playsom.html>

[6L] <http://www.cs.waikato.ac.nz/ml/weka/>

[7L] <http://www.muscle-noe.org>

Appendice: Grafici di simulazioni su un segnale musicale in alcune sottobande

