



## Messa in onda e loudness.

a cura del dott. Simone Corelli  
(GTCS, AES, Post in Europe, AIFM, ATIC)

28 giugno 2008

### Sommario

Questo documento<sup>1</sup> descrive sinteticamente gli esiti di una tavola rotonda via e-mail avvenuta tra il 27 novembre 2006 e il 13 maggio 2008 a proposito delle problematiche della messa in onda di film e di audiovisivi in genere, pubblicità comprese. Il discorso è ancora aperto ma riteniamo sia il caso di porre un... punto e virgola documentato e disponibile pubblicamente.

Tale discussione è stata voluta e organizzata dal Gruppo Tematico per la Cinematografia Sonora di AES Italia sotto la guida di Simone Corelli ed è stata particolarmente intensa tra quest'ultimo, Alessandro Travaglini, Senior Sound Designer per FOX Channels Italy e Broadcast Audio Consultant per SKY Italia, e Mauro Falcone, ricercatore presso la Fondazione Ugo Bordoni; è stata altresì seguita almeno in parte anche dal professor Aurelio Uncini de La Sapienza, Carlo Perretta di Mediaset, Gianluca Specchia di Sky, Gianni Monciotti, Alberto Colajacomo, Fabio Felici, Gilberto Martinelli, Angelo Bonanni, Maurizio Argentieri di AITS, Giovanni Belletti della RAI.

Sarebbe certamente stata utile una maggior partecipazione da parte di RAI e Mediaset che invitiamo ad intervenire al più presto, insieme a tutti gli altri broadcaster italiani.

<sup>1</sup>Il nome del file è "MessainondaGTCS.pdf" e può essere richiesto scrivendo al Gruppo Tematico per la Cinematografia Sonora all'indirizzo di posta elettronica [info@gtcs.it](mailto:info@gtcs.it) o collegandosi al sito web [www.gtcs.it](http://www.gtcs.it).

### 1 Sintesi

Il problema principale riscontrato nella messa in onda televisiva di materiale di varia provenienza è relativo alle variazioni di "loudness", ovvero d'intensità percepita, di volume, anche tra stazione televisiva e stazione televisiva, rendendo necessario per lo spettatore il ricorso al telecomando, con relativa distrazione, disappunto, scomodità, possibilità di disturbare, specialmente nelle ore tardo-serali o notturne qualora si passi ad un canale dal volume notevolmente più alto e non si provveda con velocità ad attenuarlo.

Esiste un preciso limite tecnico (EBU R-68) che fissa un tetto per i picchi del materiale audio consegnato per la messa in onda, pari a non più di 9 dB (misura quasi-peak) rispetto al livello di allineamento, in genere identificato da un tono sinusoidale a 1000 Hz di picco digitale pari a  $-18$  dB FS.

Ciò non risulta sufficiente per evitare che i pubblicitari consegnino materiale particolarmente aggressivo e disturbante, complici i sempre più potenti compressori multibanda disponibili sul mercato, che riescono ad aumentare il livello RMS mantenendo invariato il picco. A titolo di curiosità si può citare l'onda quadra a circa 3'500 Hz con picco al suddetto valore di  $-9$  dB FS come il segnale più intenso che si possa trasmettere pur rimanendo nei limiti di picco imposti.

D'altra parte ammesso di trovare un modo per valutare il livello percepito — il loudness appunto — in modo affidabile e conveniente (molto valido ad esempio il recente algoritmo proposto nel docu-

mento ITU-R BS 1770 ([8]), che ha introdotto le LU, *Loudness Units*), ciò non garantirebbe affatto un allineamento di programmi differenti portandoli al *giusto livello*: perché mai un concerto d'arpa dovrebbe suonare come una pubblicità che inizi con due trombe in primissimo piano o come un film d'azione molto intenso e rumoroso?! Pareggiandone il loudness medio, senza quindi prenderne in considerazione la tipologia, il livello di un dialogo normale risulterebbe, nel caso di un presentatore nel concerto d'arpa, molto più alto di quello del film d'azione, essendo rispettivamente il dialogo più alto della media nel caso del concerto d'arpa, e più basso della media nel caso del film d'azione.

Sembrerebbe perciò che sia necessario identificare il livello del parlato, ed utilizzarlo come riferimento per allineare i vari programmi (Dolby Dialnorm). E' ovvio che l'estrema dinamica della voce non permette affidabilità assoluta, specie in programmi particolari o di breve durata: tra un bisbiglio ed un grido, specie considerando il tipo di inquadatura (primissimo piano, campo lungo), esiste una differenza di livello di alcune decine di decibel; tuttavia statisticamente il parlato, alla lunga e quindi certamente su un film, risulta abbastanza stabile, ed equivalente, nel punto d'ascolto, a circa 70 dB SPL.

Risolto il non facile compito di identificare automaticamente il parlato, operazione di cui sembra capace ad esempio l'LM100 della Dolby, resta da capire se l'allineamento a livello realisticamente proporzionato sia sufficiente a garantire un ascolto soddisfacente per lo spettatore televisivo medio; probabilmente e purtroppo no: un'ampia e realistica dinamica, ascoltando a volume più basso del dovuto e con un sistema d'ascolto inadeguato che distorce "presto", ossia proprio quel che succede per la maggior parte dei casi d'ascolto televisivo rappresentando insomma il collo di bottiglia qualitativo, risulta in tal caso dannosissima.

Anche qui Dolby ha proposto un intelligente sistema di riduzione della dinamica automatico, centrato attorno al livello del parlato, che potrebbe risolvere buona parte del problema; ma come al solito i sistemi automatici non sono (ancora) lontanamente paragonabili a quanto possa fare un essere umano esperto, e in più se si potesse lavorare ad un mixage dedicato alla fruizione televisiva (si veda certamente [2]) già poche ore dopo la realizzazione del mixage cinematografico e con lo stesso

personale tecnico si avrebbero enormi vantaggi in termini di gradi di libertà: già la sola separazione di dialoghi, rumori, musica e riverberi, trattabili in maniera differenziata, è di indubbia utilità per ottenere un risultato ottimizzato ed efficace. Si potrebbe procedere ad una misura delle prestazioni medie degli apparati televisivi, e definire un "televisore di riferimento" da utilizzare in tutte le sale mix. In verità bisognerebbe distinguere il caso dell'ascolto da televisore da quello dell'ascolto da sistema home-theater.

Per "Pompei", di Paolo Poeti, lo scrivente ha realizzato oltre al mix per il mercato anglofono e per quello nostrano, anche il relativo mix televisivo agendo sugli stems separati con una combinazione di equalizzazione dinamica, aggiunta di armoniche per simulare basse frequenze su sistemi poco dotati sotto quell'aspetto, e di correzione manuale. Sulla base di questa istruttiva esperienza sul campo egli può dichiarare che il costo di un mix, anzi di un semplice adattamento del mix, dedicato alla fruizione televisiva "tranquilla" è stimabile in circa due turni in sala mix, ossia per un costo di circa 1000 euro, che a fronte del costo totale della post-produzione di un qualsiasi film non rappresenta probabilmente un reale problema.

## 2 Estratti dalla tavola rotonda

Per approfondire e documentare quanto sopra esposto, riteniamo utile riportare stralci della tavola rotonda sull'argomento, ossia parti di e-mail tra i partecipanti. Quasi tutte le e-mail, pressoché da subito, sono state nominate in oggetto nel seguente modo: "AES 2 MO YYMMDDhhmmss" seguito eventualmente da uno spazio e da alcune parole per riassumere sinteticamente il merito dell'intervento. Con YYMMDDhhmmss si intendono 12 cifre che rappresentino nell'ordine l'anno, il mese, il giorno, l'ora, i minuti e i secondi della spedizione dell'e-mail. In verità all'inizio l'argomento non era la Messa in Onda (MO appunto), ma la Presa Diretta (PD), e perciò alcune email risultano con titoli misti.

Per iniziare ecco l'e-mail del 27 novembre 2008, di Simone Corelli, dal titolo "AES 1 PD nonché AES 2 MO: tre zone":

E' utile scomporre l'intervallo dinamico coperto dalla colonna sonora in tre zone:

- 1) quella interessata dal segnale indispensabile per la narrazione;
- 2) quella interessata dal segnale non strettamente indispensabile, seppur utile, e immediatamente sottostante alla zona 1;
- 3) quella posta sotto alla zona 2, includendo quindi il silenzio e il rumore (ossia segnale indesiderato) di fondo di varia origine, che non subiscono particolare danno se si effettua una riduzione della dinamica a loro spese, o vengono posti sotto la soglia di udibilita' da un abbassamento del livello d'ascolto o da un previsto rumore di disturbo sul lato fruitore (le case di oggi sono rumorose). Anzi potremmo dire che giovano di una certa attenuazione.

Potremmo forse chiamare queste tre bande rispettivamente zona indispensabile, zona di arricchimento, zona di fondo o silenzio (però non dimentichiamo che contiene le code dei riverberi e quindi delicata informazione sull'acustica dei luoghi).

La prima importante osservazione quando si calcola il livello medio di una colonna sonora è che non bisogna tener conto della zona di silenzio, e per la messa in onda, o più in generale per la fruizione, su sistemi d'ascolto di bassa qualità, va garantita la comprensibilità a tutti della sola zona indispensabile. La zona di arricchimento è certamente gradito sia resa udibile, ma qui va fatto un discorso di qualità: rendere questa zona forzatamente udibile a chi ascolta ad un livello estremamente basso con un sistema d'ascolto pessimo richiede per forza di cose una modifica (ossia distorsione) del mix originale che lo allontana dall'optimum riducendone proporzionalmente la fruibilità su sistemi di buona qualità e ad un livello corretto.

E' una questione di scelte e compromessi che va lasciata alla sensibilità del fonico di mix originale. Solo sulla zona indispensabile è lecito che le stazioni di messa in onda pretendano la totale comprensibilità anche a basso volume d'ascolto (-15 dB) e con

estensione in frequenza ridotta (200 Hz-8 kHz).

Chi decide cosa debba rientrare nella zona indispensabile e cosa no? Regista, montatore scena e fonico di mix. E' assurdo che il bisbiglio di decine di persone in un bar, pur comprensibile in una sala mix di qualità, vada faticosamente e forzatamente trattato in modo che sia comprensibile anche in situazioni non ottimali se il significato narrativo di quella scena è molto semplicemente che "gente chiacchiera in un bar rumoroso".

Si osservi che quando la produzione di un audiovisivo è fin dall'inizio dedicata ad un sistema a bassa qualità è consigliabile attuare strategie per ridurre la dinamica della zona 1, e preferibilmente anche della zona 2, già in fase di ripresa e montaggio della scena: se qualcuno grida lo si riprenda in campo lungo, se qualcuno bisbiglia lo si riprenda in primissimo piano. Ovviamente il tutto compatibilmente con le esigenze narrative ed espressive. Potremmo chiamarla riduzione NATURALE della dinamica, a monte.

In seguito Alessandro Travaglini ci fornì copia del documento con cui SKY definisce le specifiche di consegna da parte di terzi. Il documento, molto interessante e valido, prodotto proprio da Travaglini, si intitola "Sky Italia - Technical specifications for external tape material", ma è anche pubblicato con altro nome ([1]):

Ecco ora una parte della risposta di Travaglini a Maurizio Argentieri:

La produzione e divulgazione di quel documento del quale sono ideatore e responsabile, è avvenuta per definire standard che potessero aiutare i nostri colleghi fonici di mix nella realizzazione di mix destinati alla diffusione broadcast, siano essi programmi, documentari, film, pubblicità, talk show, clip musicali, etc etc. . . L'obiettivo è avere trasmissioni audio percepite a volume costante ed eliminare l'uso di compressori sulla catena di trasmissione, che come sappiamo possono veramente modificare il mix prodotto in sala.

Devi sapere che molti programmi vengono mixati frettolosamente, senza molta cura per i livelli di mastering.

Come ha giustamente precisato Simone, non esisteva prima nessuna guida, se non la generica indicazione di non superare un determinato livello di picco, che affrontasse il problema della percettibilità dei mix, o che distinguesse il livello giusto riservato ad un film o ad una pubblicità, che come saprai presentano caratteristiche dinamiche molto differenti.

Ebbene, l'innovazione introdotta in quel documento consiste proprio nel differenziare categorie di contenuti ed indicare per ognuna di esse, o meglio, a seconda dell'escursione dinamica che presenta, dei margini entro i quali quel mix verrà percepito a volume costante, ovvero senza indurre il telespettatore ad impugnare il telecomando ed agire drasticamente sul volume del tv (Vedi pubblicità). In allegato trovi anche la delibera della Agenzia delle Comunicazioni che OBBLIGA le emittenti radiotelevisive a trasmettere pubblicità allo stesso volume dei programmi. Per far ciò, quindi bisogna definire dei margini, altrimenti si naviga nell'anarchia più totale.

Altre gravi problematiche che ho riscontrato durante la mia lunga analisi, derivano, per esempio, dai mix di documentari. In molte case di doppiaggio sembra consuetudine tenere basso il livello della colonna internazionale e non prestare molta attenzione all'andamento della voce narrante. Capisci bene che non è questo il giusto modo di creare un mix valido e costante<sup>2</sup>. Da qui la necessità di regolamentare il prodotto indicando dei margini minimi e massimi.

...

Ho altresì suddiviso il concetto di segnale in tre categorie: picco, loudness generale e loudness dei dialoghi.

E per ogni tipologia di contenuto ho indicato dei valori minimi e massimi da soddi-

sfare. Ho poi sottolineato che questi limiti in realtà sono dipendenti dalla compressione del segnale audio (vedi differenza tra film e interstitial). Il documento citato, che vi allego, è ora utilizzato in tutte le sedi SKY e da tutte le case di produzione ed edizione che forniscono materiale per la messa in onda a tutti i canali presenti sulla piattaforma. Altri grandi broadcasters internazionali (DISCOVERY, BBC) utilizzano specifiche audio basate sugli stessi principi. Vi posso garantire che l'utilizzo delle nuove specifiche ha migliorato enormemente la qualità audio della programmazione e tutelano il lavoro di chi ha missato bene in sala.

Sempre Travaglini continua in altra email, in questo caso "re: AES 1 PD 061128142935 (anche AES 2 MO)":

Confermo che purtroppo ho potuto constatare che in molti casi i mix vengono eseguiti frettolosamente, come per moltissimi documentari doppiati (non mi riferisco a produzioni di film per cinema o per TV) ed in altri casi l'unico obiettivo è quello di schiacciare tutta la dinamica il più possibile in modo da innalzare il volume del segnale, a discapito della qualità (pubblicità). Tutto ciò genera livelli di trasmissione molto distanti tra loro e difficile comprensione di alcuni tratti di dialoghi.

Purtroppo l'utilizzo del Dialnorm oltre che essere condizionato da alcuni fattori (inserimento nei programmi dei Metadata, capacità del sistema di trasmissione di leggerli, precisione nell'interpretazione delle misure ed inserimento dei valori) non garantisce il contenimento dei livelli entro margini di sicurezza, in modo da evitare i problemi susposti. In poche parole, un mix fatto male non verrà corretto dal Dialnorm.

Nell'adattamento sonoro per TV, un bisbiglio ed un grido si differenziano soprattutto per timbro, più che per livello (nonostante ci sia una componente di volume, che però non deve essere predominante tanto da indurre il telespettatore ad alzare il volume quando c'è un bisbiglio o ad abbassarlo quando

<sup>2</sup>Nota di Simone: qui però si va ben oltre il difetto tecnico che tra l'altro non risulterà identificabile con misure automatiche sul risultato finito.

c'è un grido, a ciò penserà il fonico di mix, tenendo presente la finalizzazione per TV).

Le tabelle indicate servono esclusivamente a questo, a dare dei margini entro i quali ci si può muovere e si ha garanzia che tutto il contenuto sonoro venga percepito senza perdita di elementi importanti.

E' per questo che esistono varie tipologie di contenuti, proprio per concedere maggior libertà ai film e meno alle pubblicità, che tendono a ridurre al massimo la dinamica per avere livelli di emissione superiori.

La delibera della AGCOM è, secondo me, lacunosa perchè non analizza i segnali audio dal punto di vista psicoacustico, ma ancora puramente elettrico. Rimane valida però l'intenzione del documento. Il caso della poesia turca e delle trombe può sempre verificarsi, ma mi sembra molto meno probabile di altri episodi a noi tristemente noti.

Il silenzio è silenzio e non va considerato silenzio nei margini descritti nelle tabelle. La parte più importante è il valore di Dialogue Loudness. Ti garantisco che purtroppo nella maggior parte dei casi i mix italiano e quelli in lingua originali sono molto differenti.

Ho detto "Un mix valido e costante", non "valido cioè costante".

Cioè ben equilibrato e che abbia un volume medio costante. 20 db di intervallo medio per diffusione broadcast credo sia un buon margine, magari andando avanti scopriamo che sarebbe meglio averne 25 o 30, tenendo sempre presente che più allarghiamo i margini più qualche furbo o pigro si sentirà autorizzato a missare frettolosamente.

Sono ben accetti suggerimenti.

A seguire, dopo osservazioni di Simone, procede con "AES 1 PD 061128211123 (anche AES 2 MO)":

OTTIMO. Mi sembra che stiamo sostanzialmente dicendo le stesse cose. Quindi, ricapitolando:

E' giusto individuare elementi sonori principali (DIALOGHI, MUSICA e SFX), secondari (altri SFX) e di sfondo (AMBIENTE)

ed attribuire ad ognuno di questi dei margini entro i quali dover far rientrare i livelli di emissione del mix. Bada bene, parlo di valori espressi in dBLeq(A) ovvero Loudness Level, non livelli elettrici del segnale. Ciò significa che a determinati intervalli di Loudness potrebbero corrispondere intervalli elettrici maggiori. Non ci dimentichiamo di questo quando analizziamo le tabelle. Introdurre l'intervento di un normalizzatore esterno tipo Dolby è inattuabile perchè richiede la modifica strutturale di tutte le aree di PlayOut delle TV d'Italia, oltre che l'equipaggiamento di apparati simili nelle sale mix di migliaia di studi. Impossibile. Le specifiche SKY sono invece si facile applicazione. Basta scaricare il plug-in gratuito di DigiDesign per ProTools SignalTools<sup>3</sup>, è tutto lì. Basta poi monitorare il livello di mix e controllare che rientri entro i valori delle tabelle.

Esiste piuttosto un altro apparecchio Dolby chiamato DP600 che entrerà in commercio all'inizio del 2007. Questo dispositivo è capace di analizzare il loudness medio di un programma audiovisivo presente nel videosever e di modificarne il livello di mastering in modo da ottenere un prefigurato target level. Questo sistema funziona solo su mix già ben equilibrati, per i quali il valore medio corrisponde comunque ad oscillazioni di livello motivate da reali esigenze estetiche e non da superficialità e approssimazione. In molti casi non è così.

Sono d'accordo che in un mondo ideale potremmo non avere limiti nell'uso della dinamica. Comunque dovremo sempre fare i conti con le esigenze di trasmissione (EBUR68) anche se devo dire che sarebbe il momento di innalzare la soglia di MPL da -9dBFS a, per esempio, -6dBFS, come è in Francia. Ciò ci permetterebbe di guadagnare 3 dB di headroom<sup>4</sup>.

<sup>3</sup>Attenzione: fornisce un risultato 3 dB più basso di molti altri strumenti di misura Leq(A) in quanto non utilizza una sinusoide a piena scala come riferimento per il 100%=0 dB come suggerisce AES.

<sup>4</sup>Nota di Simone Corelli: In verità la misura quasi-peak che andrebbe adottata fornisce già circa 3 dB di meno del reale picco misurato a livello di campione digitale, e inoltre

NON VOGLIO ASSOLUTAMENTE limitare i fonici di pubblicità. Anzi, li voglio stimolare a produrre mix decenti, con la giusta escursione dinamica e la necessaria qualità. Ma purtroppo, dobbiamo tutti constatare che l'attitudine dei pubblicitari non è dare priorità alla qualità, ma preoccuparsi esclusivamente di "suonare forte".

Analizzando numerosi mix di pubblicità ho avuto conferma che la dinamica utilizzata è spesso volte di pochi dB, con dialoghi che suonano quasi telefonici. Come è scritto bene nel documento, invece, è vero che si individuano dei margini più ristretti per le pubblicità, ma è anche spiegato bene che più è ampia la dinamica del segnale più ampi sono i margini tollerati. Al contrario più è compressa la dinamica, più stretti sono i limiti (nella parte alta) richiesti. Sta al fonico di mix spingere ad un uso più esteso della dinamica, e combattere la tendenza opposta. Infatti, condivido la tua speranza di poter avere audio di qualità nei film così come nelle pubblicità.

La realtà attuale non spinge in questa direzione. Appena i pubblicitari avranno capito che non gli conviene schiacciare la dinamica perchè non ottengono il risultato sperato di avere più volume ma semplicemente perdono qualità, allora si che cominceranno a produrre pubblicità con mix di qualità superiore, e forse un giorno paragonabile a quella cinematografica.

Secondo me dovrebbero essere questi i prossimi passi:

- partendo dal documento SKY, perfezionare i valori delle tabelle
- esplicitare — come hai suggerito — che gli ambienti (silenziosi) esulano dai limiti imposti
- innalzare i limiti della R68 (a -6dBFS?)
- stimolare i pubblicitari ad utilizzare maggiore escursione dinamica

---

sono tollerati esplicitamente 3 dB ulteriori per errori umani. Infine: tra un campione digitale e il successivo, il segnale può risultare notevolmente più alto di entrambi, addirittura di un valore infinito.

Una volta fatto tutto ciò, e applicate capillarmente le specifiche finali, potremmo ottenere la rimozione di processori di dinamica nei TX ed essere (quasi) certi che i nostri mix arriveranno nelle case dei telespettatori così come li abbiamo prodotti in sala mix.

Simone Corelli propone quindi di misurare cosa succede nelle famiglie quando si guarda la tv, con alcuni esperimenti da effettuarsi ovviamente a campione. Così giustifica tale necessità nell'email "AES 2 MO 061206233123":

Se si verifica che il livello d'ascolto è, nelle situazioni normali dei film, 15 dB più basso del previsto in sala mix, si può ad esempio fissare la giusta correzione timbrica compensativa (enfasi su bassi e acuti). Si capisce anche da che livello in giù si perderanno informazioni sonore per loro discesa sotto la soglia di rumore o di udibilità. Credo proprio sia importante scoprire questo dato... no?

A seguire Corelli ha elaborato alcune osservazioni sulla differenza tra fruizione di un film al cinema e a casa (ben trattato in [2]), proposte nell'email "AES 2 MO 061214233950":

"Mix di musica per l'ascolto stereofonico casalingo e mix della stessa musica per uso cinematografico: analisi delle differenze e discussione sul modo di affrontarle."

I sistemi d'ascolto e le abitudini d'ascolto relative alla fruizione di musica stereo in ambiente domestico rispetto alla fruizione della stessa musica come colonna sonora di un film visto e ascoltato in sala cinematografica sono diversi per più motivi:

1. La dinamica e l'estensione in frequenza sulle basse frequenze nel primo caso sono tipicamente inferiori.
2. L'angolo coperto frontalmente dal primo caso è di 60 gradi contro i tipici 45 del cinematografico.
3. La posizione d'ascolto prevista nel primo caso è centrale, mentre nel secondo caso è normale che buona parte degli

ascoltatori siano fuori asse anche notevolmente, potendo quindi dichiarare che mentre l'ascolto stereo classico sfruttando la psicoacustica può generare un'ottima illusione di provenienza dei suoni da qualsiasi punto interno all'angolo di 60 gradi poco sopra citato, nel caso cinematografico le provenienze robuste sono da soli tre punti precisi: centrale, sinistro e destro. Spostamenti (panpot) tra i tre punti sono talora possibili, ma rapidi, anche per evitare udibili problemi di phasing.

4. Il livello d'ascolto casalingo è a discrezione dell'ascoltatore e spesso limitato dalla bassa qualità del sistema di diffusione, spingendo ad un ascolto circa 15 dB più basso di quello previsto dal fonico di mix all'origine. Al contrario le sale cinematografiche hanno un livello d'ascolto standardizzato, identico a quello riprodotto nel punto dove lavora il fonico di mix.

...

5. La timbrica dei due sistemi è differente per definizione: l'equalizzazione cinematografica definita dalla cosiddetta X-curve, standard ufficiale, prevede un calo sugli acuti da tenere in buon conto nell'adattare mix per un sistema per l'ascolto sull'altro sistema. Di solito basta una correzione di un paio di decibels sopra i 2 KHz.

Detto questo è già ovvio che il mixage per i due sistemi debba essere differente, ottimizzato per ciascuno di essi. Risulta ancora più d'obbligo se si considera che la musica in un film interagisce con gli altri elementi che ne compongono la colonna sonora, dialoghi ed effetti, spesso interferendo anche negativamente sulla comprensibilità del parlato.

Ecco quindi nascere l'esigenza di poter controllare indipendentemente almeno melodia ed accompagnamento, in genere ponendo la melodia (canto, o strumento solista) su una pista apposita, di solito destinata formalmente al canale centrale, per poterne

calibrare meglio l'interazione con i dialoghi, senza generare effetto pompaggio sulla base musicale che l'accompagna. Risulterà anche possibile, in casi estremi, bucare tale melodia sulle frequenze particolarmente in competizione con la comprensibilità del parlato, ossia sulle componenti tipiche delle consonanti, nonché zona di massima efficienza dell'orecchio umano, ossia attorno ai 3500 Hz.

Il discorso non si conclude qui, ma almeno un'idea speriamo di averla data.

Gilberto Martinelli il 2 gennaio 2007 ci scrive in e-mail nominata "AES 2 MO":

Auguro buon anno a tutti. Con l'occasione, una riflessione:

Ho ascoltato il messaggio del Presidente della Repubblica. Un bell'esercizio per capire i problemi della messa in onda.

In diretta tra Rai 1, Rai 2, Rai 3, Canale 5, LA7.

Beh, se passavate da un canale all'altro avreste notato con evidenza che:

**Rai 1:** Spettro piuttosto completo ma con dinamica molto compressa.

**Rai 2:** Ottimo il range di frequenza ricco di basse e poco compresso (il migliore).

**Rai 3:** Tagliato sulle basse ma con poca compressione.

**Canale 5:** Tagliato sopra e sotto dinamica compressa anche se meno di RAI 2.

**LA7:** Filtro tagliabasso come una scure, mal compresso, distorsione leggera, insomma il peggiore.

Il giorno dopo il concerto di capodanno su RAI 1, meglio stendere un velo pietoso, meglio in RF chiaro che sul digitale terrestre.

...

Saluti

Corelli a seguire scrive nell'e-mail "AES 2 MO 070111000600":

Egregi, buon anno!

Il tempo a disposizione durante le festività mi ha permesso di ripensare al problema che stiamo trattando, ovvero alla destinazione televisiva dell'audio, e di approcciarlo secondo un'ottica che ritengo più corretta. Vado quindi ad esporVi il mio pensiero.

Innanzitutto va chiarito una volta per tutte il concetto di missaggio ad un certo livello di pressione acustica: normalmente le possibilità sono racchiuse tra due estremi, rappresentati

1. dallo standard cinematografico (rp-200 smpte), che prevede che un rumore rosa di livello pari (secondo VU-meter) a quello di una sinusoide a 1 KHz il cui picco sia -20 dB FS nel dominio digitale faccia misurare, per ciascun canale frontale escluso l'LFE, 85 dB SPL (pes. C) nel punto di ascolto (ossia di mix), 3 dB in meno per i canali surround, 10 dB in più (nella banda di funzionamento prevista ossia sotto i 120 Hz) per il canale LFE. Il tutto, non dimentichiamolo, su un sistema equalizzato secondo la curva X;
2. dal tipico livello broadcast di 79 dB SPL (pes. C) (non è uno standard vero e proprio e i valori riscontrabili sono anche superiori, o lievemente inferiori come nel caso della tv nazionale belga VRT che ha scelto 77 dB SPL (pes. C)) per un rumore rosa di livello pari (secondo misuratore RMS) a quello di una sinusoide ad 1 KHz il cui picco sia pari all'allineamento in uso, ossia -18.06 dB FS (circa, in realtà un ottavo del massimo livello codificabile) in Europa, e -20 dB FS negli Stati Uniti. Esiste anche uno standard lievemente diverso, definito in Giappone se non erro, per l'HDTV, che si rifà al livello di pressione acustica totale raggiungibile con l'insieme dei canali a disposizione, ma in sostanza non ci si discosta usualmente dal valore che ho indicato.

Riguardo alla banda occupata dal rumore rosa di cui ai due punti precedenti rimando per ora la trattazione.

Si noti però che per destinazione broadcast è quasi sempre richiesta una limitazione del livello affinché non superi di oltre 9 dB il livello di allineamento fissato, misurando con un quasi peak meter (PPM), riducendo quindi la dinamica possibile e il massimo livello di pressione raggiungibile, già ridotta dall'allineamento indicato.

Osserviamo che anche l'allineamento cinematografico, apparentemente (e anche per esperienza di spettatori) capace di ricreare pressioni acustiche notevolissime, non è però sufficiente, a meno di artifici poco eleganti, a garantire la straordinaria dinamica dei suoni reali, dal rumore di due piatti che si urtano in cucina all'urlo di un uomo in primo piano, non volendo citare casi più estremi come un'esplosione o un semplice sparo in scena di guerra, la fanfara dei Bersaglieri, un piatto infranto contro un muro, il lancio di un missile, l'esplosione di una città colpita dagli alieni. E' stimabile che sia necessario aggiungere quindi una dozzina di dB di headroom al sonoro di domani, per un maggiore realismo sui picchi.

Tornando ai problemi di oggi, esiste, come dicevo, una certa confusione sull'allineamento al mix e non è raro sentir chiedere a che livello si stia missando, e quindi alla risposta 85 dB sentir commentare "che alto!". Si confonde insomma l'allineamento col livello che si decide di raggiungere: missare a 85 dB (come meglio descritto sopra) non significa che il suono che ci raggiungerà sarà a 85 dB: significa che applicando un ben determinato segnale test, come già spiegato, si otterrà quella pressione acustica; si è fissata insomma una scala, un rapporto causa-effetto.

Una normale voce narrante sarà quasi certamente ad una settantina di dB SPL, i bisbigli 10 dB più bassi e così via, e ciò sia che si missi secondo un certo allineamento che secondo un cert'altro: l'importante è che il livello di mix e quello di riascolto siano



identici. Tra l'altro, a parità di livello di riascolto, materiale missato a 85 dB suonerà più debole di quello missato a 79, come dovrebbe essere ovvio per chiunque si occupi di suono e/o abbia un minimo di capacità di ragionamento.

Consigliare dunque ad un fonico di mix di allinearsi ad un livello domestico non significa affatto indurlo a generare un mix che suoni più debole! Il fonico cercherà sempre di missare ricreando il livello naturalmente suggerito dall'immagine.

Dunque si tratterebbe invece di riascoltare il proprio normale mix al livello probabilmente scelto dal fruitore televisivo medio e verificare se in seguito a tale attenuazione (di 15 dB normalmente) vada perduta la comprensibilità di qualche passaggio narrativamente importante, o se l'apparente calo di bassi e acuti dovuto alle ben note non linearità dell'orecchio umano sia tale da suggerire una compensazione.

E' anche prevedibile che l'ascoltatore medio mal tollererà livelli superiori a 75 dB SPL (C) se non per brevissimi istanti, per vari motivi (inclusa la bassa qualità ossia la grande distorsione introdotta dal sistema d'ascolto medio televisivo) e che scomparirà sotto la soglia del rumore di fondo del mondo moderno molto materiale di arricchimento quali gli ambienti più delicati (stormire di fronde, cinguettii, avvicinarsi di un temporale, chiacchiericcio in strade lontane et cetera) e le code di riverberazione, con conseguente perdita di informazione sullo spazio architettonico che ospita le vicende. Anche qui è consigliabile un'azione tecnica di recupero, rialzando questi ingredienti sonori sopra la prevedibile tipica curva di rumore o soglia di udibilità.

Ripetiamo: lavorare già al livello di fruizione medio porterà a rialzare tutto il materiale, probabilmente schiacciandolo contro la soglia imposta per materiale broadcast e quindi introducendo con i limiters distorsione non lineare... indi l'ascoltatore riabbasserà ulteriormente il livello d'ascolto! Si lavori a dinamica e livelli naturali, e poi si effettuino manualmente o in automatico (vedremo

come) le correzioni necessarie per rendere accettabile un ascolto con le caratteristiche peggiorative che abbiamo illustrato.

Proprio questo argomento va ora affrontato, ma è necessaria un'indagine seria per misurare in modo affidabile:

1. Qual'è il livello d'ascolto e l'intervallo dinamico graditi dal fruitore televisivo medio (e ci piacerebbe capire anche il perché, pur avendo buone ipotesi di risposta al quesito);
2. Qual'è la banda di frequenze che ci dobbiamo aspettare realisticamente. . . anzi: qual'è in generale la risposta in frequenza che dobbiamo aspettarci, facendo anche attenzione alla variabilità nella banda e alla fastidiosità di picchi/valli nella risposta (infatti basta un picco di risposta a causa di un'onda stazionaria ad esempio a 400 Hz per spingere ad abbassare il livello d'ascolto per non essere troppo disturbati.  
Sulla base di ciò se, proseguendo in questo esempio, scopriremo che il 20% dei televisori presentano tale difetto mentre il rimanente 80% lo compensasse saremmo indotti ad attenuare questa frequenza anche se la media dei televisori mostrasse una risposta - MEDIA APPUNTO - lineare);
3. Qual'è la distorsione tipicamente introdotta e secondo quali leggi si genera.

E qui mi fermo in attesa della Vostra collaborazione: va organizzato un test serio su almeno una dozzina di casi reali.

Io posso mettere a disposizione materiale audiovisivo test, ben conoscendo il livello d'ascolto ritenuto naturale in fase di mix avendo io stesso missato tale materiale. Uomini di buona volontà fatevi avanti!

...

Travaglini risponde in "re: AES 2 MO 070111000600":

Sono pienamente in accordo con quanto scrivi.

Vorrei solamente suggerire di non tendere ad appiattire la qualità dei mix al fine di renderli fruibili necessariamente anche ai sistemi di riproduzione più scadenti, considerando anche il fatto che inevitabilmente il mercato tende ad innalzare la qualità dei prodotti in vendita (vedesi l'aumento di TV stereo e di sistemi Home-Theatre, anche se spesso di qualità medio-bassa). Sarebbe inoltre penalizzante per chi ha scelto di godere appieno del contenuto sonoro trasmesso.

Per ciò che riguarda l'intervallo di dinamica tollerato in ambiente domestico suggerirei di applicare le indicazioni ([11]) dell'Ing. Thomas Lund della T.C. Electronic, [http://www.tcelectronic.com/media/lund\\_2006\\_loudness\\_nab2006.pdf](http://www.tcelectronic.com/media/lund_2006_loudness_nab2006.pdf) mentre riproporrei le mie tabelle, ovviamente modificabili e migliorabili se necessario, per ciò che concerne i limiti minimi e massimi per categorie sonore (dialoghi, effetti, fondi, ecc.). Non lo dico per vanagloria ma perchè frutto di un lungo lavoro di analisi che ci può far giungere ad un risultato positivo in breve tempo.

In "AES 2 MO 070112222538" Corelli scrive:

Note sparse:

1. L'ascolto casalingo di materiale di varia provenienza tramite sistema televisivo o home theater risulta effettuato da buona parte dell'utenza, e in special modo per il materiale tipicamente broadcast come telegiornali, pubblicità, programmi di varietà, soap, ad un livello di ascolto più basso di quello realistico (ossia previsto da un buon fonico di mix su sistemi ad alta dinamica come quelli cinematografici), ad un livello insomma che potremmo chiamare *comfortevole* in quanto permette di alternare visione+ascolto a chiacchiere con i familiari, altre attività, telefonate anche solo spostandosi di pochi metri o nella stanza attigua. . .
2. I limiti dell'ascolto casalingo tipico se presi troppo in considerazione portano a mix che poi si riveleranno di bassa qualità una volta che il prodotto venga ascoltato con sistemi buoni, e considerando il probabile miglioramento (tv digitale, multicanale) futuro dei sistemi casalinghi, porterebbe ad un'occasione persa nelle repliche della messa in onda. Meglio creare un buon mix e poi ridurlo e ottimizzarlo per le varie destinazioni in modo preferibilmente automatico (quindi con una riduzione di costo), archiviando comunque il mix originale di buona qualità per il futuro.
3. Tra i limiti tipici dell'ascolto casalingo permette anche agli altri abitanti dell'appartamento di proseguire le proprie attività senza essere troppo distratti o disturbati.

Altro motivo per un livello d'ascolto basso è rappresentato dalla bassa qualità dei sistemi di riproduzione MEDI presenti nelle case, specie amplificazione+casse incorporate nei televisori, anche in quelli di ultima generazione. Ciò induce appunto ad un livello basso per non essere troppo disturbati da distorsioni varie. Va incluso nel discorso anche l'ambiente d'ascolto che è sempre affetto dal problema delle onde stazionarie, anche molto intense, generando picchi nella risposta in frequenza che inducono ad abbassare l'intero programma (è il motivo per cui in cuffia si ascolta ad un livello decisamente superiore mancando questo problema).

A tale diversità di livello non si può reagire abbassando il livello di calibrazione nella sala mix, in quanto tale operazione indurrebbe il fonico ad incidere ancora più forte. . . Il livello finale *comfortevole* va testato dopo aver effettuato il mix, ed applicando quindi azioni d'ottimizzazione che vedremo in seguito di consigliare. Tra l'altro buona parte del materiale che va in onda è di derivazione cinematografica. . .

go va considerata la scarsa estensione soprattutto sulle basse frequenze (circa 150 Hz invece dei 16 Hz necessari ad esempio per la nota più bassa che può emettere un organo a canne, assolutamente udibile pressoché da chiunque dal vivo e di grande impatto emozionale). Questo limite va considerato per gli elementi narrativi fondamentali (terremoto in un film...) e risolto in questo caso introducendo artificialmente componenti sonore di frequenza più alta, ad esempio con tecniche di simulazione percettiva (introduzione di armoniche che suggeriscano al cervello l'esistenza di una fondamentale che invece non si ode) (maxxbass...).

4. Il livello d'ascolto basso (circa -15 dB, da verificare con esperimenti al più presto) implica la necessità di compensare la timbrica (effetto loudness) con parametri da studiare, purtroppo variabili col livello istantaneo del programma in ascolto (differenza tra *equal loudness contour* originale e quello del segnale attenuato a livello casalingo).
5. Chi deve effettuare l'adattamento? Il fonico originale o la stazione di messa in onda, o ancor meglio il dispositivo ricevente se è in grado di ricevere qualità alta da rielaborare? Si potrebbe pensare anche ad una trasmissione in parallelo di due versioni, tra le quali ogni utente possa scegliere.
6. Rimane da considerare la differenza angolare sulla stereofonia, e sul campo visivo abbracciato ([2])!

Il professor Uncini può inviarcì qualche tesi-sta per acquisire dati utili a risolvere questi punti?

E ancora, nel tentativo di riassumere il tutto in "AES 2 MO 070205212109 dolby volume":

Ritengo che possiamo cominciare a scrivere, in attesa di qualche volenteroso che faccia per noi alcuni test... per ora lasceremo in bianco alcuni dati.

Ecco:

Esigenze di ordine pratico per la messa in onda di materiale di varia provenienza da parte di una stazione televisiva, e comodità per l'utente televisivo stesso affinché non debba scegliere un volume d'ascolto diverso per ogni stazione trasmittente, e per ogni programma, suggeriscono certamente una standardizzazione tecnica, ovvero un allineamento tra segnale, in forma analogica (elettrico, magnetico...) o digitale che sia, e pressione acustica da generare nel punto d'ascolto. Sarebbe altresì importante standardizzare almeno i parametri fondamentali che caratterizzano la qualità d'ascolto: estensione in frequenza, curva di risposta, rapporto tra segnale diretto e riverberazione tipica che realisticamente ci si deve aspettare dal locale in cui si ascolta, massima potenza indistorta, livello di rumore di fondo, posizione geometrica dei diffusori attorno all'ascoltatore, dimensione e posizione dello schermo contenente l'immagine... in quanto condizionante anche l'audio.

Il mondo del cinema ha fissato le sue regole da tempo.

Il televisivo si è dimostrato invece meno serio, fissando in modo troppo semplicistico limiti di picco nel livello nella modulazione, ma non dichiarando esplicitamente null'altro di quanto sopra esposto, o facendolo senza essere ascoltato. In particolare il livello d'ascolto di riferimento non è definito, e varia in genere da 77 a 87 dB SPL (pesati C) per canale, alimentando con un rumore rosa limitato in banda audio e di livello RMS (o vumeter?) pari a quello della nota di allineamento a 1 KHz, a sua volta pari a -18.06 dB FS di picco nel digitale per il broadcast europeo, -20 dB FS in quello di scuola statunitense. 10 dB di variabilità sono ovviamente troppi...

inoltre la disponibilità di compressori/limitatori multibanda di grande potenza, ha portato in questi ultimi 10 anni ad un incremento del livello medio rispetto al livello di picco permesso, e quindi risulta che gli spot pubblicitari, o la musica, modulano talmente "hot" da risultare in un

livello ottimale d'ascolto ancora inferiore, ossia circa 70 dB SPL (C) sullo stimolo di allineamento poc'anzi citato (e conseguente riduzione della dinamica).

La variabilità di livello quindi, già alla fonte, raggiunge il ragguardevole valore di 17 dB. Segue poi la trasmissione (specie analogica), che porta ad ulteriore variabilità da stazione televisiva a stazione televisiva.

In ogni caso pur risolvendo il problema con una standardizzazione seria del livello, rimane da affrontare il problema della scelta a gusto del livello d'ascolto da parte del telespettatore. Nella maggior parte dei casi le cause per un ascolto a livello inferiore di quello corretto, deciso dal fonico di mix in condizioni controllate, sono due:

- Scarsa qualità del sistema d'ascolto, che porta a non tollerare troppa distorsione, e quindi a mantenere il livello basso.
- Desiderio di non disturbare vicini o famigliari, di poter in ogni momento conversare con i famigliari o permettere loro di farlo senza essere troppo disturbati dal suono proveniente dal televisore... ciò porta al concetto di *volume d'ascolto confortevole in condizioni casalinghe di convivenza con altri, e con sé stessi* ossia all'ascolto DI SOTTOFONDO, non disturbante, ma pur sempre tale da permettere la comprensione dell'evento, ossia soprattutto la comprensione dei dialoghi. Si parla di circa 55 dB SPL (C) in media invece di circa 75 (misurerò meglio domani).

...

da cui deriva un documento dal nome "Livelli080207.pdf" che viene inviato agli altri ne "AES 2 MO 070215205510 testo bozza iniziale".

Poi in "aes 2 mo 070421003920" sempre Corelli scrive:

Se si decide che il loudness medio deve stare in certi limiti, preferibilmente bassi, si permette un guadagno di dinamica, e si evita che i pubblicitari facciano spiattellare il mix

verso l'alto con conseguente riduzione dei picchi rispetto al valor medio. Certo! Sarebbe un bene per il suono. Un grande passo avanti.

Ma, sempre rimanendo nell'ambito della pubblicità, rimane da valutare il rapporto tra loudness misurato, e loudness giusto, naturale: perché una pubblicità bisbigliata e delicata musicalmente dovrebbe essere mandata in onda allo stesso livello di una pubblicità dove suona una banda militare con trombe in primo piano? Tarando il livello d'ascolto nella propria abitazione in modo che sia adeguato per la banda militare, ipotizzando che vada in onda per prima, ascolteremo la pubblicità bisbigliata ad un livello UGUALE a quello della banda, ossia infinitamente più alto di quel che sarebbe NATURALE udire. Ovviamente ho fatto un esempio estremo.

Purtroppo il livello NATURALE lo può conoscere solo un essere umano, non una macchina che non distingue tra grida, bisbigli, voce che esce da un telefonino o da un'ugola in primo piano et cetera. E' quel che fa il fonico di mix: decide il livello naturale, giusto, adatto. Se lo fa in un ambiente controllato, con un guadagno d'amplificazione standardizzato e fisso, come succede nel cinema, possiamo mandare in onda quel materiale senza problemi: sarà tutta NATURALE, da film a film, da spot a spot. Se invece lo fa un fonico televisivo, dove non esiste un vero e proprio standard ma solo abitudini approssimative, è il disastro: materiale imprevedibile, da fonico a fonico, da anno ad anno...

Basterebbe fissare il rapporto segnale digitale vs. pressione acustica come nel cinema! Poi se al controllo tecnico Mediaset o chi volete voi ascolta il materiale, con lo stesso rapporto nel sistema di riproduzione, e scopre che è AGGRESSIVO, FASTIDIOSAMENTE sproporzionato e innaturale, lo rimanda indietro con giustificazione come la seguente: Siamo spiacenti di comunicarLe che lo spot da Lei consegnato, ascoltato al livello previsto dal documento tal dei tali, firmato da tutte le emittenti nazionali, risulta innaturalmente troppo intenso, di

TOT decibels, infatti la voce, decisa ma non gridata, si situa ad un livello di TOT decibels sopra a quanto indicato nel suddetto documento.

Insomma questo documento dovrebbe dire, per esempio:

- voce narrante: -25 dB FS, misura Leq(A)
- grido di persona in primo piano: max XXX dB FS

Insomma vanno fissati, semplicemente con una giornata di lavori in cui invitare 10 fonici di mix d'esperienza, parametri, almeno per la VOCE, che farà da appiglio, da riferimento, per proporzionare gli altri ingredienti sonori.

Lo so, si possono fare esempi estremi, ipotizzare difficoltà varie... ma sarebbe già un inizio. Anche perché in Italia le pubblicità le missano POCHISSIMI studi. Basta andare là con un fonometro e un rumore rosa test, e regolarli una volta per tutte l'amplificazione secondo quanto si deciderà (diciamo di seguire il cinema, il che semplifica tante cose!). Dopodiché egli saprà che il controllo qualità userà lo stesso standard, e se missa roba troppo, troppo, troppo aggressiva, gli torna indietro con scritto AGGRESSIVA.

Scusate se ho scritto di getto, malamente. Prendeteli per appunti da sgrossare. Attendiamo che Travaglini termini il suo impegno in AES a Vienna e poi riprendiamo il discorso.

Travaglini in "aes 2 mo 070422225153":

In un contesto di ascolto televisivo in ambiente domestico, i margini tra livelli minimi e livelli massimi sono molto ridotti.

A causa delle esigenze di trasmissione, è sensato ipotizzare uno scenario in cui anche le sonorità più estreme (come quelle portate ad esempio da Simone) verranno comunque riprodotte a livelli di loudness che non avranno differenze superiori ai 10dB Leq(A). Dalle mie ricerche il livello medio, pari a

-25dBLeq(A), è un'ottimo punto di riferimento. Su questo valore si può standardizzare il livello di riproduzione dei dialoghi (è tarato così il canale di informazione SKY TG24). In termini di livelli di picco, generalmente qualsiasi tipo di programma audio non raggiunge i -10dBFS. Pertanto gli attuali limiti pari a QPPM -9dBFS (True Peak -6dB) sono più che sufficienti per qualsiasi esigenza produttiva.

Rimane il problema della normalizzazione del Loudness per segnali codificati con Metadata, per i quali gli attuali decoder automaticamente riportano l'emissione a -31dBLeq(A). La Dolby sta lavorando a modificare questo parametro in modo da guadagnare qualche dB.

Il problema con i pubblicitari è facilmente risolvibile. Basta richiedere la masterizzazione degli spot a specifici livelli di loudness. Nel caso SKY Italia, ad esempio, pubblicità e promo devono avere loudness tra -25 e -23 dBLeq(A).

Ci sono ottimi segnali positivi che alimentano la mia fiducia per il prossimo futuro: SKY Italia e FOX Channels Italy hanno inserito le specifiche Loudness in tutti i contratti con i fornitori di materiale audiovisivo. Altre emittenti televisive in onda sulla piattaforma (Universal, Jimmy, Alice, Nuvolari, etc...) hanno investito, o stanno pianificando di farlo, risorse per equipaggiarsi con strumenti di misura e processori di loudness adeguati, in modo da allinearsi con gli altri canali del gruppo. Molte società audiovisive si stanno attrezzando con loudness meter e stanno cercando di capire di cosa si sta parlando e come ciò potrà modificare il loro prodotto.

Travaglini poi, a seguito di osservazioni di Corelli, risponde in "aes 2 mo 070423160502":

Rispondendo ai tuoi quesiti ecco le risposte.

1. Le esigenze televisive probabilmente sono differenti da quelle cinematografiche e non richiedono una dinamica

così ampia. La consuetudine di utilizzare limitatori sul master permette di disporre di headroom più che sufficiente.

2. Tutti i ricevitori satellitari sono in grado di decodificare Metadata e tra questi il parametro Dialnorm (Dialogue Normalization). Per impostazione di fabbrica, al momento NON modificabile, il segnale audio viene normalizzato ad un valore equivalente ai -31dBLeq(A). Ovvero, se il Dialnorm è -25, il mix audio viene ridotto di 6 dB. Se il Dialnorm è -29 viene ridotto di 2 dB, e così via. In questo modo tutti i programmi vengono riprodotti allo stesso valore di loudness equivalente, pari a -31 dB. Come detto, vari broadcasters stanno segnalando alla Dolby che questo genera discontinuità di volume tra programmi con Metadata e programmi senza (questi ultimi generalmente hanno loudness pari a -25): Pertanto la Dolby sta lavorando alla possibilità di introdurre una regolazione di guadagno sui decoder domestici in modo da compensare queste differenze.
3. Nelle specifiche SKY Italia, si richiede di masterizzare pubblicità e promo con loudness Short Term incluso tra -25 e -23 Leq(A). E' fondamentale ricordare che lo Short Term misura gli ultimi 10 secondi del segnale, e di questi dà una media, aggiornando continuamente la sua misura. Oltre alle misurazioni Loudness, le specifiche impongono di contenere anche i livelli elettrici entro precisi margini, rispettando i quali è impossibile fare ciò che ipotizzavi nella tua mail. Sono comunque consentite escursioni di volume, a patto che rientrino nei margini stabiliti. Escludo comunque uno scenario come quello descritto, dato che il primo e solo obiettivo dei pubblicitari è quello di vendere bene il loro prodotto nel brevissimo tempo messo a loro disposizione.

Successivamente, dopo una fase di studio di un

documento fornitoci da Mauro Falcone, si pensa di analizzare statisticamente la distribuzione dei livelli in un film, e Corelli in "aes 2 mo 070605210230" scrive:

Riguardo al problema della finestra (e limitiamoci a quella rettangolare, perché ci sarebbe da discutere anche su questo) della misura RMS (o altro tipo) per ottenere il grafico della distribuzione dei livelli di un programma audio, ho pensato a lungo. Per dimostrare quanto sia importante e influente sulla forma del grafico che si ottiene, considerate ad esempio i seguenti casi:

- Una sinusoide a 20 Hz, a livello  $y$ , udibilissima se il sistema d'ascolto lo permette, con una finestra ben superiore a 50 millisecondi mi porterà ad un istogramma che indica appunto una componente sola a livello  $y$  e basta. Bene. Se la finestra è inferiore avremo un disastro, ovvero tutt'altra distribuzione!
- Se la finestra di tempo è pari all'intero programma abbiamo un solo valore, ossia una distribuzione piatta di livello pari all'RMS medio dell'intero programma. Viceversa se la finestra di tempo è grande quanto un singolo campione, abbiamo una distribuzione molto variegata anche con una sola sinusoide.

Quel che dobbiamo chiederci è: qual'è la finestra giusta per i nostri scopi? Sicuramente superiore a 50-70 ms per evitare che componenti audio possibilissime (come un DO basso di un organo a canne (16 Hz)) creino problemi.

Scendiamo più in profondità... a noi serve una finestra tale che renda MOLTO simili i grafici di programmi audio di varia provenienza, a meno di una traslazione, che è il valore di normalizzazione che cerchiamo!

E il Travaglini pare averlo trovato, con finestra appunto ESAGERATA, pari all'intero programma, e pari a -25 dB FS. Tuttavia la media RMS su tutto il programma, pesato

A, vale per buona parte dei programmi audio, ma non per quelli un po' estremi... tipo film MOLTO AGITATI.

Penso che si potrebbe perfezionare il tutto se prelevassimo solo il canale centrale, eventualmente estraendolo con dolby decoder da segnali stereo LR, perché esclude buona parte della musica e un po' di effetti (tipo terremoti o roba imprevedibile che va sull'LFE). Insomma contiene quasi soprattutto dialoghi e quindi è un canale maggiormente utile come riferimento per dedurre il livello a cui deve stare il programma.

Allora urge un test: prendere una decina di film, tutti missati di norma nelle stesse condizioni di livello e di timbrica, cosa verificabile ascoltandoli ad orecchio, e verificare se hanno tutti lo stesso valore RMS medio pesato A sul canale CENTRALE. In tal caso prendere tale tecnica come valida procedura normalizzante.

Poi bisogna adattare i programmi alla dinamica del mezzo, e all'abitudine ad ascoltare a 15 dB di attenuazione. Ovvero bisogna applicare la compressione sui segnali forti e un innalzamento sui segnali deboli. Cosa che Dolby prevede e rende possibile nei lettori DVD. Più un po' di effetto LOUDNESS gonfiando soprattutto i bassi.

Ok, erano un po' di ragionamenti post cenam... appunti al volo da discutere e su cui ragionare meglio.

Mauro Falcone ci informa riguardo agli studi di AGCOM per il loudness televisivo in "R: aes 2 mo 070726103000 delibera AGCOM su volume TV":

Gentilissimi Colleghi,

vi informo che l'Agenzia per le garanzie nelle Comunicazioni (AGCOM), ha finalmente deliberato relativamente al problema del volume televisivo delle pubblicità, instaurando un tavolo tecnico sull'argomento.

L'Agenzia tipicamente contatta le direzioni delle Emittenti che conseguentemente instaurano internamente le procedure di adesione. Tuttavia credo sia bene da parte di tutti voi, seguire questa cosa nell'interesse

di tutta la comunità tecnica scientifica, e spero di trovarvi numerosi a contribuire in tale ambito istituzionale. nel testo della delibera trovate tutte le indicazioni e le persone da contattare in AGCOM per eventuali informazioni.

Vi comunico infine che la delibera AGCOM è stata pubblicata sulla Gazzetta Ufficiale GU n.170 del 24 Luglio 2007, e che quindi i 30 giorni citati nella delibera scattano da tale data.

Anche sulla base di questo interesse vorrei rilanciare l'idea di organizzare un workshop sull'argomento in AES ed eventualmente estenderlo ad altre associazioni interessate. Ovviamente a maggior ragione rilancio la volontà di organizzare un incontro informale tra tutti noi per parlare un poco a ruota libera sul problema e su come affrontarlo.

Vi auguro un felice periodo di vacanza, o di lavoro per chi avesse già usufruito o avesse spostato in tarda estate il meritato riposo.

Qui trovate la delibera in questione, sito AGCOM: [http://www.agcom.it/provv/d\\_50\\_07\\_CSP.htm](http://www.agcom.it/provv/d_50_07_CSP.htm)

...

Se volete, per vostra convenienza, qui trovate un mio zip con tutti i documenti AGCOM e della GU già organizzati. <http://www.fub.it/voice/temp/AGCOMSPOT.zip>

Il 13 maggio 2008 le comunicazioni (per ora) si concludono, con una segnalazione da parte di Simone del sito web <http://www.orban.com/meter/> dove sono presenti softwares di misura (pare stia per rendersi disponibile anche un Dolby Digital Meter, versione software del LM100, disponibile sia in stand-alone che come plug-in RTAS, e va citato anche TC Electronic LM5, plug-in Pro Tools TDM) ed alcuni interessanti documenti sulla problematica (<http://www.orban.com/support/orban/techtomics/>). In uno di questi leggiamo:

1. L'attuale tecnologia per il controllo automatico del loudness può controllare efficacemente tale parametro, ma introduce inevitabilmente come effetto secondario una notevole limitazione della dinamica.

2. Per il materiale ben missato si consiglia di disinsere il controllo automatico di loudness. Altro materiale, come le news o i programmi sportivi, non possono essere curati non avendo tempo a disposizione, e perciò un controllo automatico del loudness è consigliabile.
3. In generale, quando è richiesto un risultato di valore, non è per ora possibile fare a meno dell'intervento umano, in quanto solo gli umani sono in grado di valutare le sottigliezze del contesto.

L'ultima osservazione suggerirebbe di studiare la possibilità di sostituire i processori di loudness con alcune persone che a turno valutino ogni programma ascoltandone alcuni punti a campione per deciderne il volume corretto. Dato che per ogni programma sono probabilmente necessari non più di 30 secondi per effettuare l'operazione, in media, si deduce che l'ipotesi non sia affatto da trascurare.

Per la prosecuzione dei lavori è necessario che un tesista con capacità informatiche ci coadiuvi sviluppando un semplice software per rilevare la distribuzione statistica di picchi, RMS e altri tipi di misura audio, con finestra d'integrazione a scelta, di files audio mono o stereo. L'utilizzo di tale software applicato a mix e premix vari porterà a conclusioni importanti.

## Riferimenti bibliografici

- [1] Alessandro Travaglini - Sky Italia, "Specifiche tecniche per materiale audio-video".
- [2] G. Martinelli, S. Corelli, "La costruzione del suono per l'immagine cinematografica: dalla presa diretta al mixage, ed il rapporto con l'ambiente home-theater", paper 05001 AES italiana, presentato al SIB di Rimini nell'ambito della conferenza "Multichannel Audio: per l'immagine e per il suono", Marzo 2006.
- [3] S. Klar, G. Spikofski, "On levelling and loudness problems at television and radio broadcast studios", AES Paper 5538 - Presentato alla 112esima Convention AES del Maggio 2002, Monaco, Germania.
- [4] N.H.C. Gilchrist, "DRACULA: Dynamic Range Control for Broadcasting and Other Applications", BBC Research and Development Report no. 1994/13 (1994).
- [5] N.A.F. Cutmore, "Dynamic range control in a multichannel environment", JAES, aprile 1998, volume 46, numero 4, pp.341-347.
- [6] Neville Thiele, "Some thoughts on the dynamics of reproduced sound", JAES, Gennaio/Febbraio 2005, volume 53, numero 1/2.
- [7] J. Couling, "TV Loudness: Time for a New Approach?".
- [8] Raccomandazione ITU-R BS.1770<sup>5</sup>, "Algorithms to measure audio programme loudness and true-peak audio level".
- [9] J. P. Moerman, "Program Loudness: Nuts & Bolts", Convention Paper 6436, presentato alla 118esima AES Convention, Maggio 2005, Barcellona, Spagna.
- [10] J. Couling e D. Pope, "Reducing Loudness Variation on Broadcast Services by Utilising Automatic Speech Detection".
- [11] T. Lund, "Control of Loudness in Digital TV".
- [12] J. C. Riedmiller, S. Lyman, C. Robinson, "Intelligent Program Loudness Measurement and Control: What Satisfies Listeners?", presentato alla 115esima Convention dell'AES - Ottobre 2003, New York, Stati Uniti d'America.
- [13] M. Falcone, A. Barone, A. Bonomi, G. Monaco, D. Ciavatta, "Abbassa quello spot, per favore!".
- [14] Alessandro Travaglini, "Broadcast Loudness: Mixing, Monitoring and Control", Convention Paper 7044, presentato alla 122esima AES Convention, Maggio 2007, Vienna, Austria.
- [15] S. Corelli, F. Felici, G. Martinelli, "Elementi di Cinematografia Sonora", Edizioni Lambda, (Italia, 2006).

<sup>5</sup>A fine Maggio il gruppo ITU che cura il tema Loudness si è riunito ed ha prodotto due nuovi draft documents, le cui versioni definitive integreranno la 1770.