

Appunti (in via di sviluppo) di tecnica del suono: livelli

dott. Simone Corelli, simonecorelli@mac.com

13 giugno 2005

Indice

1 Livelli audio	1
1.1 Definizioni	1
1.2 Livello di allineamento elettrico	2
1.3 Livello di allineamento digitale	3
1.4 Livello di allineamento acustico	3
1.5 9 dB di headroom per destinazioni broadcast	4
2 Misuratori di livello audio	4
3 Toni di allineamento	6
4 Osservazioni sul livello di registrazione dal vivo	6
4.1 nuovo livello di riferimento (-32 dB FS) per il <i>nuovo audio</i>	7
A Il decibel	8

1 Livelli audio

1.1 Definizioni

- Il *livello massimo permesso* (PML=Permitted Maximum Level) è il livello che viene superato solo raramente dai picchi del *programma*. Con un misuratore quasi-peak (si veda la def. a pag.5) non va superato. Con misuratori di picco vero è consentito un margine di circa 3 dB oltre questa soglia. Esiste un'ulteriore tolleranza di 3 dB per errori umani.
- Il livello di allineamento audio (AL=Alignment Level)¹ è il livello di un segnale sinusoidale ad 1 kHz attenuato di 9 dB rispetto al *livello massimo permesso*.

¹definito nella raccomandazione ITU-R BS.645-2: *Test signals and metering to be used on international soundprogramme connections* e ripreso ad esempio nelle raccomandazioni EBU R73-1999

- Il *livello di misura*, ML, viene fissato a 12 dB sotto al livello di allineamento. A questo livello di sicurezza vanno inviati segnali per la misura di risposta in frequenza come spazzolate sinusoidali.

Se monitorati su un misuratore di livello di tipo *ppm* (Peak Programme Meter, si veda la definizione a pag.5) i programmi più comuni² moduleranno tra 6 e 9 dB oltre il livello di allineamento, mentre con un VU meter (si veda la definizione a pag.5) la lettura si muoverà sostanzialmente attorno a 2 dB sotto al livello di allineamento.

- Il *livello standard operativo*, SOL, è definito come il valore da non superare con segnali stazionari, e dato che sono esplicitamente permessi picchi che eccedono tale valore di 8 o 9 dB si deduce una certa somiglianza con il *livello di allineamento* che però si differenzia in quanto superabile anche da segnali stazionari. Dunque $ML < AL \leq SOL < PML$.
- Il *livello di riferimento* è indicato come livello 0 dB o anche 100%. In certi casi viene invece definito come massimo livello ammesso! MOLTO IMPORTANTE CHIARIRE!!!!
- Il *livello test*, o nota di controllo. viene fissato dall'UER a 0 dBu, e pari a 9 dB sotto al livello massimo. Se ne deduce la corrispondenza con il livello di allineamento.
- L'*headroom* è... (definito dopo. Riordinare la posizione della definizione mettendola solo qua)

1.2 Livello di allineamento elettrico

Il livello di riferimento (in seguito ribattezzato e ridefinito come *livello di allineamento*) elettrico per l'audio è nato nell'industria telefonica ed è stato fissato dai laboratori Bell nel 1939 pari ad una tensione tale da dissipare 1 mW (da qui la lettera emme) su un carico tipico di 600 Ω. Tale valore elettrico si indica con 0 dBm (600 Ω) ed è quindi pari, per una sinusoide, a $0.775 V_{RMS}$ (RMS: *root mean square*, in italiano *valore efficace*), approssimazione ormai universalmente adottata (l'errore è inferiore a mezzo centesimo di decibel) di $\sqrt{\frac{600}{1000}}$ V; la tensione *picco-picco* è invece $0.775 \cdot \sqrt{2} = 2,1920310 \dots$. La Bell contemporaneamente introdusse il VU Meter (VU=Volume Unit, pari a 1 dB), strumento visivo atto a verificare il corretto allineamento elettrico (si veda l'approfondimento a pag.5).

In seguito nel Nord America venne adottato un *livello operativo standard* (l'acronimo in lingua inglese è *SOL*) pari a +8 dBm, definendolo come valore da non superare con segnali stazionari. Tuttavia alcune autorità e molti stabilimenti al loro interno adottarono +4 dBm, oggi forse il valore più diffuso, equivalente ad una tensione RMS di $10^{\frac{4}{20}} \cdot 0.775 \text{ V} = 1.2282922 \dots \text{ V}$ pari (per una sinusoide) ad un valore *picco-picco* di 3.4741350... V.

Data la successiva obsolescenza tecnica del valore di impedenza di 600Ω si passò in seguito ad indicare la tensione di riferimento di 0.775 V direttamente con 0 dBu e non più con 0 dBm, rendendola indipendente dall'impedenza.

²Decisamente difficile accordarsi stabilmente su cosa si intenda esattamente con questa definizione che quasi invita a produrre programmi con poca dinamica.

1.3 Livello di allineamento digitale

Il riferimento per il dB FS (che misura quindi per definizione 0 dB FS) è una sinusoide di frequenza 997 Hz, il cui picco positivo corrisponde al massimo consentito nel dominio digitale (0FFF in esadecimale, per una quantizzazione a 16 bit, e 0FFFFFF a 24 bit, mentre il picco negativo vale F000 e F00000 rispettivamente). Si preferisce usare tale frequenza al posto di 1000 Hz per evitare di usare solo 48 campioni dei 2^{numbit} disponibili utilizzando la diffusissima frequenza di campionamento di 48000 campioni al secondo.

Esistono principalmente due standard di allineamento tra segnali digitali ed elettrici. Quello secondo le norme EBU R68 e ITU richiede che un tono sinusoidale (preferibilmente di frequenza 997 Hz) di ampiezza di picco -18.06 dB FS circa (per la precisione un ottavo del livello massimo (*Full Scale*, appunto), quindi $20 \log_{10} \frac{1}{8} = -18.0617997\dots$) debba valere 0 dBu (misura RMS); quello invece fissato dalle norme americane *SMPTE RP155* richiede che il suddetto tono sinusoidale abbia valore di picco -20 dB nel dominio digitale e +4 dBu nel dominio elettrico. La differenza tra i due standard è dunque di 6 dB (quasi il 200%) a favore dell'SMPTE.

1.4 Livello di allineamento acustico

L'allineamento acustico professionale più in uso consiste nell'utilizzare un *rumore rosa* la cui ampiezza RMS sia la stessa di una sinusoide di frequenza 997 Hz e ampiezza di picco -18 dB FS, regolando il sistema d'ascolto per ottenere una pressione acustica di 85 dB SPL (pesati C) per ogni diffusore acustico³ (corrisponde grosso modo ad un *mezzoforte* orchestrale ascoltato al centro di un auditorium); nello standard cinematografico i due canali surround vengono però settati 3 dB più bassi. Il canale LFE (*Low Frequency Effect*) infine viene settato affinché nella sua banda (limitata a 150 Hz, o 120 Hz secondo le più recenti indicazioni EBU) sia, in ogni terzo d'ottava considerato, 10 dB più alta della risposta nella banda media degli altri diffusori frontali. Attenzione: nello standard cinematografico si usa una curva di equalizzazione del sistema di ascolto che attenua notevolmente gli acuti (da approfondire). Osserviamo anche che a causa del problema di onde stazionarie in ambienti di dimensioni domestiche o anche più grandi, nonché per la differenza di estensione (soprattutto alle basse frequenze) esibita da diversi diffusori acustici, risulta consigliabile l'utilizzo di un rumore filtrato, che presenti un massimo di energia sonora attorno ai 600 Hz, al posto del ben più critico rumore rosa. Si noti che il livello d'ascolto domestico risulta in genere 6 dB più basso di quanto qui descritto (quindi 79 dB SPL (C)), obbligando a compensare con 6 dB di guadagno sul materiale registrato. E' per questo motivo che nella raccomandazione tecnica R68-2000 la EBU consiglia di trasferire materiale digitale *broadcast* (viene citato in particolare il DAT) che sia quindi allineato con AL=-18 dB FS e PML=-9 dB FS amplificandolo di 6 dB se destinato a riproduttori CD AUDIO casalinghi tarati per l'ascolto di CD commerciali; ciò restringe l'headroom dai 9 dB *broadcast* ai 3 dB tipici del mercato *home*. Va verificato con attenzione che questa operazione non dia luogo a clipping. Terminiamo con l'osservazione che questo allineamento non permette una dinamica particolarmente ampia, infatti la riproduzione di un semplice parlato di una singola persona alla distanza apparente di un metro e mezzo circa

³in alcuni testi si trova -20 dB FS e 83 dB SPL, che è esattamente la stessa cosa.

fa raggiungere su un singolo canale picchi a ben -6 dB FS circa. Probabilmente tra qualche anno questo allineamento verrà quindi modificato.

1.5 9 dB di headroom per destinazioni broadcast

In ambito *broadcast* esistono limiti molto severi riguardo al livello da non superare, per evitare ad esempio sovramodulazioni nei trasmettitori FM, saturazioni sui nastri magnetici analogici (max 510 nWb/m) o di superare la massima deviazione (+/- 50 kHz) nel sistema PAL. Dunque esiste una regola fondamentale per destinazioni broadcast: non superare di oltre 9 dB (misura *quasi-peak*, che in genere equivale a circa 3 dB in meno di una vera misura di picco⁴) il livello di allineamento. Per completezza osserviamo che alcune organizzazioni preferiscono invece indicare il limite di 8 dB.

Vanno utilizzati a tale scopo compressori/limitatori opportunamente calibrati, oppure, volendo evitare una manipolazione della dinamica del programma, va ritarato il livello di allineamento 0 VU, ponendolo pari a 9 dB in meno del picco massimo raggiunto dal programma audio, chiaramente se il rapporto segnale/rumore del mezzo di destinazione permette una buona intelligibilità nei passaggi più delicati. Questa non è tuttavia una soluzione elegante in quanto comporta la perdita del riferimento d'ascolto fisso, a meno di non allineare definitivamente 0 VU pari a -9 dB FS (in RAI si usano i -10 dB FS), essendo garantita l'impossibilità di superare lo 0 dB FS di picco⁵.

La questione rimane aperta. In ogni caso gli standard di allineamento tra digitale e videoregistratori analogici Betacam SP più in uso in postproduzione sono, dal più frequente al più raro:

- -14 dB FS
- -18 dB FS o simile, secondo EBU R68 e ITU
- -20 dB FS, secondo SMPTE RP155
- -10 dB FS, in uso presso RAI

ma è ovvio che se non ci si attiene esplicitamente all'headroom raccomandato, e ad un livello d'ascolto standard, l'allineamento vede ridotta la sua utilità.

2 Misuratori di livello audio

Gli strumenti visivi di misura del livello audio si distinguono per il tipo di media effettuata e per il loro comportamento in regime dinamico (tempo di reazione e di rilascio). I più raffinati fanno uso di conoscenze di psicoacustica per visualizzare il livello in modo particolarmente conforme alla pressione sonora percepita (*Loudness*).

Fondamentalmente possiamo distinguere:

⁴e come se non bastasse vengono tollerati altri 3 dB autorizzando, se così si può dire, picchi (brevi) ben 15 dB sopra al livello di allineamento. Si veda a tal proposito il documento *EBU Technical Recommendation R68-2000* a pag.1

⁵in realtà è possibile che il segnale convertito in analogico, tra un campione e il successivo, superi tale valore.

1. misura della media del valore assoluto
2. misura del *valore efficace*, ossia RMS (radice quadrata del segnale elevato al quadrato)
3. misura VU Meter (1940) intermedia tra le due precedenti e con 300 millisecondi di tempo d'integrazione ossia circa la durata di una sillaba. In base al documento *A New Standard Volume Indicator and Reference Level, Proceedings of the I.R.E., Gennaio 1940*, il VU meter originariamente utilizzava rettificatori ad onda intera all'ossido di rame che, combinati con lo smorzamento elettrico, effettuavano una media definita da una risposta secondo la formula $i = k \cdot e^p$ equivalente alla risposta effettiva dello strumento per *deflessioni normali* (approfondire) (Nell'equazione i è la corrente istantanea nella bobina dello strumento ed e è il potenziale istantaneo applicato all'indicatore di volume). In un gran numero di VU meter è stato rilevato un valore p pari a circa 1.2 ossia un comportamento intermedio tra la misurazione media ($p = 1$) ed RMS ($p = 2$). La scala del VU meter si estende da -20 a +3 dB indicati anche con valori percentuali e lo 0 VU=100% può essere settato secondo il livello operativo in uso (di solito +4 dBu o +8 dBu); le sue caratteristiche dinamiche sono tali da far raggiungere alla lancetta la deflessione del 71% dopo 0.3 s dall'azzeramento del segnale d'ingresso, caratteristica scelta in base alla durata media di una sillaba ma che rende possibile che transitino picchi musicali o di parlato anche 10 dB superiori all'indicazione fornita dall'apparecchio. Ciò ha portato a costruire apparecchi in grado di reggere queste brevi sovramodulazioni introducendo al massimo l'1% di *distorsione armonica totale* (l'acronimo in lingua inglese è *THD*) e alla misura di tale tolleranza dinamica è stato dato il nome di *headroom*. In genere l'headroom varia da 10 a 24 dB a seconda della qualità del dispositivo.
4. misura PPM (acronimo di Peak Programme Meter), anche definita *quasi-peak*, ossia di picco sulla media del segnale con tempo d'integrazione breve, pari a 5 (tipo I) o 10 millisecondi (tipo II), basati sul fatto che picchi più brevi di questo valore possono essere limitati senza conseguenze udibili
5. misura di picco, ossia istantanea su ogni singolo campione; nel caso di una frequenza di campionamento di 48000 campioni al secondo ciò significa avere un tempo d'integrazione di circa 20 microsecondi. E' una misura utile per evitare *clipping*, ma poco rappresentativa del *loudness*. Osserviamo che raggiungere gli 0 dB FS non è garanzia certa di clipping, sebbene una sequenza di almeno 3 campioni a questo livello ne siano indicazione altamente probabile. Sulla sua udibilità maggiore o minore con materiale sonoro particolare si può disquisire a lungo ma, nel dubbio, è meglio farne a meno dato che al contrario di certe altre distorsioni (come quelle introdotte dalle valvole) il clipping di certo non è mai gradevole.

Sulla base di queste categorie principali si sono sviluppati indicatori di livello ben definiti, con scale in dB, numerazione, allineamento, comportamento dinamico ed escursione specifici. I più usati sono:

1. Nordic N9

2. Din PPM
3. BBC PPM

DA COMPLETARE anche con tabella e foto.

3 Toni di allineamento

Una pratica utile e necessaria per garantire la correttezza degli eventuali trasferimenti di materiale audio è quella di registrare in testa al programma toni di allineamento che identifichino i canali e il livello di allineamento (di conseguenza si definiscono automaticamente il PML e il ML). Nel caso di materiale mono si utilizza un tono sinusoidale a 997 Hz (o 1 KHz), nel caso stereofonico il tono assegnato al canale sinistro (canale 1) è bucatato ogni 3 s per un tempo di 250 ms, pratica utile anche per identificare un mix monofonico proveniente da mix stereofonico. Utile aggiungere che per materiale multilingua si usa scegliere i 400 Hz per identificare la lingua secondaria. Come già detto il livello di allineamento nel dominio digitale vale -18.06 dB FS di picco, come da raccomandazioni EBU R68-2000, indipendentemente dalla risoluzione in bit del segnale. Per materiale multicanale (intendiamo con più di due canali) l'EBU descrive una struttura di questo genere:

1. su tutti i canali 3 secondi di tono sinusoidale coerente (in fase) alla frequenza di 1 KHz, tranne l'eventuale LFE cui è assegnata la frequenza di 80 Hz
2. per ogni canale, tranne l'LFE, partendo dal canale frontale sinistro e proseguendo in senso orario, mezzo secondo di pausa e mezzo secondo di tono
3. ulteriori 3 secondi di tono coerente su tutti i canali
4. il canale LFE proseguirà invece senza pause con il tono a 80 Hz. Si noti che di solito la riproduzione di questo canale è amplificata di 10 dB, ma la sensibilità dell'orecchio compenserà questo fatto facendo percepire un'intensità molto simile a quella degli altri canali

Osserviamo uno dei vantaggi di questa metodologia: qualsiasi normale submix sarà in grado di mostrarci come è stato realizzato, e il livello originario di ogni canale.

4 Osservazioni sul livello di registrazione dal vivo

Con l'avvento della quantizzazione a 24 bit, e quindi di un incremento (teorico) della dinamica di quasi 50 dB rispetto all'uso di 16 bit (per la precisione $20 \log_{10} \frac{2^{24}}{2^{16}} = 48.1647993$ dB, si rende possibile utilizzare un livello di registrazione fisso, standardizzato, basato per comodità sull'allineamento acustico di cui sopra (sezione 1.4, a pag. 3) o più basso, per garantire un margine di sicurezza atto ad evitare clipping nelle situazioni più a rischio; osserviamo infatti che

- La posizione del microfono può non riflettere la posizione virtuale dell'ascoltatore scelta successivamente, in fase di missaggio. Ad esempio: in genere il microfono sul set di un film viene piazzato a poche decine di centimetri dalla bocca dell'attore mentre l'inquadratura potrebbe suggerire una distanza dell'ascoltatore pari ad un paio di metri.
- E' raro (purtroppo) che si voglia riprodurre un suono con una pressione sonora realistica. Tipico il caso di uno sparo, di un'esplosione, ma anche di un grido o di una porta che sbatta violentemente. Sono tutti casi che, realisticamente, richiederebbero valori SPL impossibili allo standard cinematografico attuale, a meno di non utilizzare trucchi come l'uso di più canali simultaneamente, l'uso del canale LFE (che, come detto, garantisce 10 dB di livello massimo in più) per le frequenze più basse, uno *scrambling* del suono per disallineare le fasi delle componenti frequenziali attenuando il picco del segnale senza ridurne il valore RMS, l'aggiunta di riverbero et cetera.

Probabilmente 14 dB in meno dello standard cinematografico si riveleranno adeguati creando però un iniziale scompiglio tra i montatori, abituati a vedere forme d'onda mediamente più gonfie e a tenere la manopola dell'amplificatore nel primo terzo della sua escursione, salvo poi dover inseguire nei vari ciak i diversi livelli di registrazione e incontrare o generare spesso saturazioni da clipping digitale. Purtroppo l'abitudine a vedere il suono invece di sentirlo, di ascoltare ad un volume inferiore a quello realistico (colpa degli appartamenti sempre più piccoli e sempre meno isolati gli uni dagli altri, del rumore di fondo che ci accompagna nelle città, del non spendere più di 500 euro per l'impianto d'ascolto) hanno portato ad una diminuzione sensibilissima della dinamica, aspetto incompreso, spesso confuso con la massima pressione acustica, maltrattata e trascurata, e invece importante corresponsabile dell'impatto emozionale del suono.

Sarebbe auspicabile un

4.1 Nuovo livello di riferimento (-32 dB FS) per il *nuovo audio*

Il miglioramento tecnologico di questi ultimi 5 anni sta rendendo sempre più diffuso l'utilizzo del campionamento a 24 bit, e di un numero di tracce superiore alle solite due per il suono in presa diretta cinematografico. Ciò richiede alcuni chiarimenti affinché questa rivoluzione qualitativa comporti il minimo sforzo per essere apprezzata completamente, senza che qualcuno vi rinunci a causa di pregiudizi infondati. Con questo documento vogliamo portare chiarezza, e renderci disponibili ad eliminare ogni eventuale dubbio residuo.

Registrare in digitale a 24 bit invece che a 16, come il DAT ci aveva costretti a fare, significa disporre di una precisione sonora 256 volte maggiore o, se vogliamo, di una gamma dinamica migliorata di oltre 48 dB per contenere contemporaneamente picchi di suono senza clipping e suoni delicatissimi perfettamente "leggibili" dall'orecchio. E' molto importante qui sottolineare che le diffusissime vecchie macchine Avid per il montaggio della scena non gestiscono i files audio a 24 bit⁶, ed è quindi spesso necessario che il fonico fornisca una versione ridotta a 16 bit, preferibilmente utilizzando una tecnica detta *dithering* che garantisce files a 16 bit di gran qualità, equivalente a circa 18 bit. Il montatore del suono poi dovrà riportare automaticamente il lavoro all'originale qualità di 24 bit (eventualmente multitraccia)

⁶le macchine recenti non hanno questo limite

utilizzando i files registrati del fonico di presa diretta (magari multicanale) al posto di quelli (magari mono) che gli fornirà l'esportazione OMF. Inizialmente di fronte a registrazioni a 24 bit correttamente realizzate o a loro riduzioni per Avid molti saranno spiazzati da un livello di registrazione mediamente molto più basso di quello a cui si era abituati. Non è un errore del fonico e non è un difetto! Semplicemente si alzi il volume d'ascolto agendo sull'apposita manopola dell'amplificatore (a questo punto sarebbe auspicabile un nuovo livello di riferimento a -32 dB FS per 85 dB SPL (pesati C), ossia che permetta 14 dB di dinamica in più, da allineare a -10 dBu invece che a +4 dBu ovviamente). L'unico vero aspetto sgradevole di tutto ciò sarà una lettura delle forme d'onda grafiche sullo schermo meno agevole (finché non si decideranno a disegnarle in scala logartimica). Il vantaggio varrà lo sforzo, senza alcun dubbio.

A Il decibel

Il decibel (dB) non è un'unità di misura: per capirci, a nessuno verrebbe in mente di definire il simbolo di percentuale (ossia %) come un'unità di misura dato che semplicemente rappresenta un modo per confrontare due valori omogenei, per l'appunto raffrontandone il rapporto al valore 100, scelto come riferimento; il rapporto tra due valori omogenei è infatti adimensionale. Allo stesso modo il *bel* e il suo sottomultiplo *decibel*, che vale un decimo di bel, non sono unità di misura ma modalità di rappresentare, in scala logaritmica, adatta quindi ai nostri sensi, anziché lineare, il rapporto tra due valori omogenei. Per l'esattezza il bel è il logaritmo in base 10 del rapporto tra un valore di potenza ed un altro di riferimento mentre il decibel, ben più utilizzato, è pari a un decimo di bel (quindi 1 bel=10 dB). Dato che la potenza è funzione del quadrato dell'intensità vale la seguente formula per il calcolo del rapporto tra due intensità, espresso in decibel: $10 \log_{10} \frac{i^2}{i_{ref}^2}$ semplificabile nella nota espressione $20 \log_{10} \frac{i}{i_{ref}}$.