

I segreti dell'MQA

MQA è noto come formato di compressione delle tracce audio ma è anche tema di discussioni incentrate sul fatto se sia un formato lossy o lossless, ovvero con o senza perdita di informazioni.

La realtà MQA è un sistema ibrido, ad alta risoluzione, 192 kHz / 24 bit, che punta a due obiettivi primari:

- elevato risparmio di spazio dati;
- miglioramento della messa a fuoco temporale delle tracce audio.

Obiettivi entrambi indirizzati non solo alle nuove registrazioni ma a tutto l'archivio delle registrazioni master disponibili.

Il presente articolo non è e non vuole essere un trattato accademico, poiché ha solo scopo divulgativo, con l'intento di rendere l'argomento comprensibile ad un ampio pubblico.

Nei fatti, MQA lavora sulla banda di frequenze da 0 Hz a 96 kHz e, come vedremo, è sostanzialmente lossless, ovvero senza perdita di dati, almeno fino a 48 kHz, mentre nell'intervallo 48-96 kHz è lossy, ossia comporta una parziale perdita di dati. Sotto questo aspetto è quindi meno prestante del formato FLAC, il lossless per eccellenza, rispetto al quale offre tuttavia una capacità di compressione tre volte superiore. Questo è un dato facile da calcolare, confrontando i due formati (vedi **Tabella 1**); più complesso è invece dare una valutazione agli effetti della perdita di informazione dell'MQA nell'intervallo 48-96 kHz; possiamo però notare, parlando di musica, che questo intervallo corrisponde a una sola "ottava" delle oltre 13 che si possono ragionevolmente prendere in considerazione come massima estensione dell'insieme delle frequenze contenute in qualsiasi registrazione di musica. Quindi, nel caso dell'MQA, possiamo considerare la copertura di



almeno 12 ottave senza perdita di informazione e "perdita" nell'ultima ottava delle 13.

In realtà i promotori di MQA sostengono che questo formato può essere considerato lossless per l'impiego audio poiché, come vedremo, utilizza un algoritmo di incapsulamento e non di compressione; in più, con il filtro di de-blurring (focalizzazione) implementato, migliora la messa a fuoco temporale delle tracce audio conferendo maggiore naturalezza in fase di riproduzione.

Per approfondimenti si riportano i seguenti link della presentazione tenuta presso la Audio Engineering Society nel 2014 e gli estremi del brevetto:

Convention Paper 9178 - Audio Engineering Society - A Hierarchical Approach to Archiving and Distribution bit.ly/2V4ePVn

Patent Scope: WO2013186561-Doubly Compatible Lossless Audio Bandwidth Extension: bit.ly/2V5JL7o

MQA - Master Quality Authenticated

L'algoritmo MQA ha due obiettivi primari:

- occupare meno spazio di memorizzazione senza compromettere le informazioni audio;
- migliorare la riproduzione audio mettendo a fuoco il segnale.

Come occupare meno spazio dati conservando le informazioni fondamentali per l'orecchio

Nel capitolo 3, "L'orecchio, il campionamento a 44,1 kHz, a 192 kHz, e 24 bit di risoluzione", sono esaminate le caratteristiche dell'orecchio umano che si possono così sintetizzare:

- lo spettro di frequenze comprese tra 20 Hz e 20 kHz è rilevato dalla coclea per la comprensione del messaggio sonoro;
- per catturare i *twig snaps*, ossia segnali brevissimi ed impulsivi, si devono distinguere segnali che distano 6 μ s, pertanto occorre campionare a 192 kHz;
- per coprire la gamma dinamica di 120 dB dell'orecchio umano si devono usare almeno 20 bit che, nel mondo dei computer, diventano 24.

I segnali campionati a 192 kHz / 24 bit soddisfano questi 3 requisiti ma, per

Tabella 1

Tipo di file	MegaByte di hard disk per minuto di musica (stereo)	Album .wav/.aiff per 1 TB	Album .flac per 1 TB
16 bit / 44,1 kHz (qualità CD)	10 MB	2.000	4.000
16 bit / 48 kHz	11 MB	1.900	3.800
24 bit / 48 kHz	16,5 MB	1.200	2.400
24 bit / 96 kHz	33 MB	630	1.260
24 bit / 192 kHz	66 MB	320	640
DXD (24 bit / 352,8 kHz)	121,12 MB	165	330
DSD64	40,375 MB	500	1.000
DSD128	80,75 MB	250	500

Nota: il numero di album per 1 TeraByte di spazio a disposizione è approssimato e basato su album di 45 minuti

contro, lo spazio occupato è oltre 6 volte maggiore rispetto a quello necessario per un normale audio CD, ossia 44,1 kHz / 16 bit.

Il formato .flac comprime il file audio ed è un formato lossless ossia privo di perdite (è l'analogo di un file .zip in informatica). A priori una stima dello spazio che occupa un file .flac rispetto, ad un .wav o .aiff, è impossibile perché la compressione, in termini informatici e non audio, dipende dalla dinamica del brano, dal formato di partenza e dallo strumento di conversione. A grandi linee però lo spazio occupato viene dimezzato.

L'altro aspetto da considerare è il tempo necessario per trasferire i file dai fornitori di servizi di streaming, Tidal, Spotify, Qobuz o qualsiasi altro, al consumatore finale per l'ascolto.

Per le reti cablate questo non è un problema dato che praticamente in tutta Italia lo streaming di video 4K è disponibile; se si ricevono file video 4K, l'audio, anche in alta risoluzione, è un "gioco da ragazzi".

Ma in un futuro molto vicino, o per le aree non ancora servite dal servizio internet cablato, vi sarà l'esigenza di trasmettere su reti radio 3G o 4G con destinatari i dispositivi mobili, si pensi ad un smartphone equipaggiato con cuffie hi-fi. In questi casi, la banda è una risorsa costosa e non infinita, e perciò la differenza è fatta dagli algoritmi di compressione.

Diventano quindi fondamentali le performance degli algoritmi di compressione. Se per analogia si guarda allo streaming video, dovrebbero essere note ai più le differenze di prestazioni dell'algoritmo di compressione di Netflix rispetto a quello di Prime Video; gli effetti sono visibili specie all'inizio dei film quando il buffer si sta caricando.

Pertanto lo spazio occupato da un file da trasmettere è un parametro fondamentale.

Tidal, Qobuz o qualsiasi altro fornitore al momento propongono il servizio di streaming ad un numero limitato di utenti dato il costo mensile non proprio per tutti. Se dovesse arrivare un terzo attore ed offrisse il servizio alta risoluzione per 3,99 euro/mese, ci si potrà aspettare un'esplosione dello streaming e quindi la necessità di una compressione efficiente per raggiungere un numero di utenti pari a 100.000, 1.000.000 di volte l'attuale. Non è banale.

Chi pensa al proprio NAS da 12 TB, contenente oltre 7.000 album/opere in 192 kHz / 24 bit in formato .flac, si riterrà al riparo dalla cosa perché lo spazio di registrazione ha oggi costi oramai irrisori. Un proprio ar-

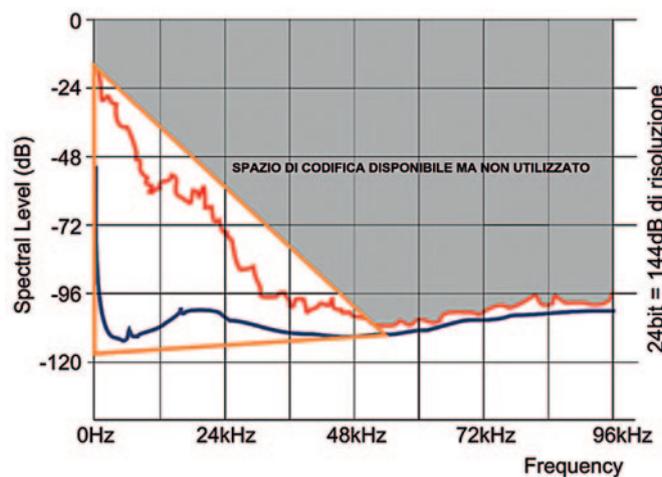


Figura 1: contenuto spettrale comune a qualsiasi brano musicale.

chivio, trascurando la questione della copia e dei diritti, non si può confrontare con un servizio di streaming che può consentire, non solo di ascoltare i brani, ma anche di accedere a tutti i contenuti connessi a ciò che si sta ascoltando: copertine, testi, booklet, storia, release. La non elevatissima diffusione in Italia dello streaming audio è al momento solo una questione di canone.

MQA Ltd., per anticipare la risposta alla necessità di ottimizzare i dati da trasmettere in rete, ha architettato un algoritmo ibrido di incapsulamento /compressione.

La ricerca ad opera di Gerzon dell'Oxford's Mathematical Institute, Graven di Algol Applications e Stuart di Meridian Audio, pubblicata in "Convention Paper 9178 - Audio Engineering Society - A Hierarchical Approach to Archiving and Distribution", ha analizzato il contenuto spettrale di musica in tutte le forme nella banda 0-96 kHz con 144 dB di dinamica; questo coincide nel mondo digitale con l'alta risoluzione, ossia 192 kHz di campionamento e 24 bit di risoluzione.

Il risultato relativo all'analisi di un particolare brano musicale si può sintetizzare con l'analisi spettrale rappresentata in **Figura 1**:

- la curva in rosso riporta i valori di picco del brano musicale alle varie frequenze;

- la curva in blu riporta il rumore medio ambientale ed elettronico.

I loro studi hanno dimostrato che il risultato è sostanzialmente lo stesso indipendentemente dal brano e dal genere musicale, ossia una curva sempre molto simile a quella rossa, con preponderanza di informazione sotto i 48 kHz e appiattimento sotto i -96 dB oltre i 60 kHz. Si può notare che lo spazio di codifica evidenziato in grigio è vuoto, ossia non vi sono mai tracce audio da campionare in quell'area; di conseguenza questo è uno spazio disponibile ma mai utilizzato.

Inoltre, si può notare che il contenuto spettrale del brano musicale e il rumore di fondo convergono, quasi si incrociano a 60 kHz, per poi proseguire appaiati fino ai 96 kHz.

Si può pertanto individuare un'area, il triangolo arancione, all'interno della quale spazierà qualsiasi brano musicale e dove avrà senso campionare a 24 bit ed operare in modalità lossless.

Oltre i 60 kHz, avere 24 bit di risoluzione, ossia 144 dB di dinamica disponibili, è uno spreco perché lì non sarà mai presente un segnale musicale. Oltre i 60 kHz vi sarà comunque del segnale utile alla corretta ricostruzione temporale del segnale, pertanto da non filtrare.

Lo studio, poi proseguito e confermato da altri ricercatori, ha dimostrato che:

1) non è sufficiente campionare a 96 kHz / 24 bit perché, anche se con dinamica ridotta, vi è del segnale utile fino ai 96 kHz; pertanto, per rilevare tali componenti del segnale, si deve campionare a 192 kHz. Si deduce quindi che l'alta risoluzione è utile e necessaria per un ascolto completo; sopra i 60 kHz si è molto vicini alla soglia del rumore e si ha una dinamica di soli 48 dB (nella Figura 1 si noti che il limite inferiore è a -144 dB ed il segnale è sempre sotto ai -96 dB, perciò $144 \text{ dB} - 96 \text{ dB} = 48 \text{ dB}$) equivalenti a 6 bit di risoluzione; la gamma dinamica della parte udibile, quella entro i 24 kHz, contenuta entro il triangolo arancione in Figura 1, può essere risolta con 18 bit e infatti la curva rossa non scende mai sotto ai -108 dB.

In sintesi, l'informazione musicale occupa solo una parte dello spazio disponibile 96 kHz / 144 dB; la prima collocata alle basse frequenze, triangolo arancione, dove si ha la massima gamma dinamica e dove bastano 18 bit dei 24 disponibili, la seconda, ultrasonica, dove bastano 6 bit per ricostruire il segnale originale.

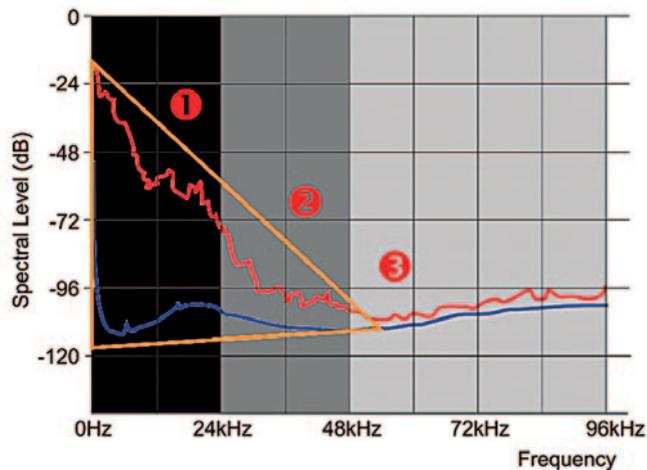


Figura 2: bande audio e ultrasoniche.

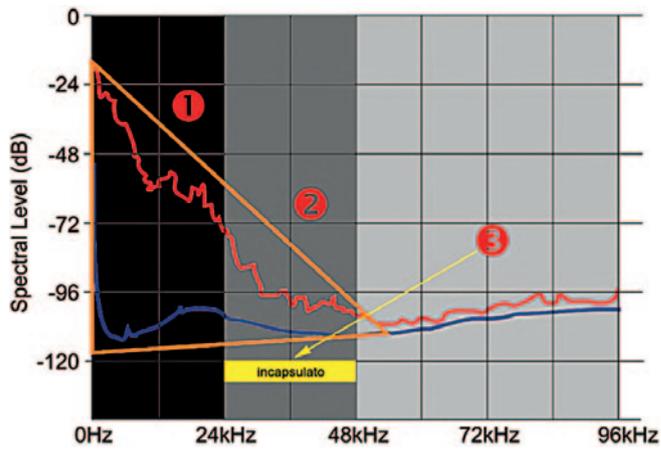


Figura 3: frequenze oltre i 48 kHz incapsulate.

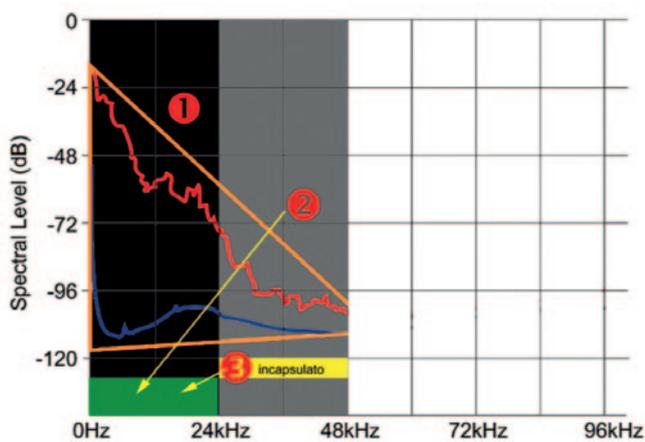


Figura 4: frequenze ultrasoniche compresse lossless.

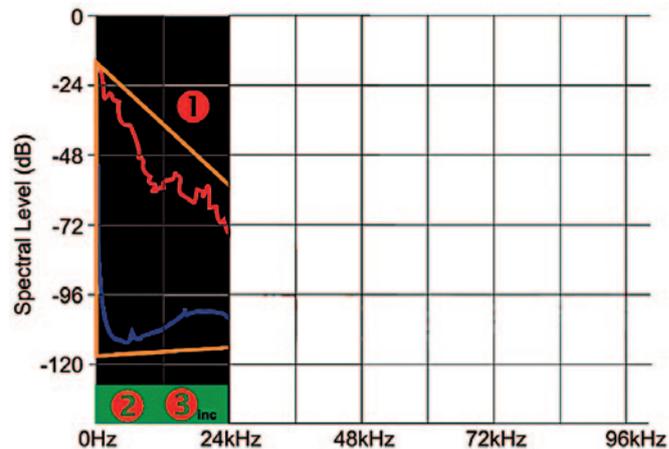


Figura 5: banda audio lossless in 18 bit.

A partire dagli studi accennati e che accomunano tutti i generi e brani musicali, MQA Ltd. ha messo a punto una tecnica di “ripiegamento” dei campioni considerando solo i bit utili e scartando tutte le parti di spettro che non conterranno mai informazioni provenienti da registrazioni audio. L’algoritmo MQA ripiega come in un origami le informazioni audio in alta frequenza in banda base, ma in un modo tale da poterle estrarre e ricostruire in un secondo momento, a livello di dispositivo di riproduzione (DAC), per una riproduzione senza perdite.

In **Figura 2** è riportato, come visto in precedenza, lo spazio spettrale, rappresentato dalla curva rossa, praticamente comune ad un qualsiasi brano musicale; sono evidenziate però le 3 zone di frequenza coinvolte nell’audio:

- le frequenze di campionamento di base, 44,1 kHz o 48 kHz, generano campioni nella zona in nero ❶;
- le frequenze di campionamento doppie, 88,2 kHz e 96 kHz, generano campioni nella zona in grigio ❷;
- le frequenze di campionamento quadruple, 176,4 kHz e 192 kHz, generano campioni nella zona in grigio chiaro ❸.

Il contenuto **percepibile** dall’orecchio umano è maggiormente concentrato in ❶, con una quantità comunque non trascurabile anche in ❷ corrispondente di fatto al triangolo arancione. Il contenuto audio nella regione ❸ è trascurabile. Qui si trova solo rumore ambientale ed elettronico utile comunque per la ricostruzione temporale del segnale e sua messa a fuoco (de-blurring del segnale).

Nelle **Figure 1, 2, 3, 4 e 5**, la scala dovrebbe essere logaritmica, ma per necessità di dettaglio, la banda ❸, 48 kHz - 96 kHz, è riportata in scala doppia rispetto a ❶ e ❷. In questo modo il confronto della curva rossa con la curva blu nella banda ❸ è evidenziato. L’algoritmo MQA si può sintetizzare in 3 step.

1° step dell’algoritmo MQA

Tutto il contenuto spettrale, curva in rosso, della regione ❸, praticamente privo di audio e contenente solo rumore ambientale ed elettronico, viene incapsulato sotto al livello di rumore (curva blu) oltre i -130 dB, ma viene comunque mantenuto distinto dal segnale in ❷.

Non si tratta di un classico downsampling; il processo, denominato di incapsulamento, è tecnicamente una forma di riduzione dei dati con perdita. Questo processo di incapsulamento con perdita, tuttavia, è considerato accettabile dai sostenitori del sistema MQA perché generalmente non ci sono informazioni audio nella regione ❶ ma solo rumore come evidenziato dagli studi citati precedentemente; si tratta quindi di essere in grado di mantenere l’alta frequenza di campionamento originale quando il segnale verrà decodificato dal dispositivo di riproduzione (DAC), al fine di una corretta ricostruzione temporale.

2° step dell’algoritmo MQA

Dato che la dinamica è entro 110 dB, tutto il contenuto spettrale della regione ❶ viene memorizzato nei 18 bit più significativi dei 24 bit del campione memorizzato, utilizzando una tecnica di “noise shaping”. Successivamente, tutto il contenuto spettrale della regione ❷, contenente anche la regione ❸ incapsulata, viene trasferito lossless sotto al livello di rumore della regione ❶ utilizzando i 6

bit meno significativi dei 24 bit del campione memorizzato.

Risultato finale

Nella regione ❶ si trovano i campioni della banda originale 0 - 24 kHz, lossless, memorizzati nei primi 18 bit; questi sono seguiti dai 6 bit che contengono la regione ❷, anch'essa mantenuta lossless, e che contiene a sua volta la regione ❸ incapsulata.

Dato che la tecnica di origami è voluta e nota, quando si tratterà, a livello di dispositivo di riproduzione (DAC), di riesplodere il segnale nella sua forma originale, si applicherà esattamente il processo inverso.

Le regioni ❶ e ❷, perciò, saranno identiche al segnale originale, lossless, mentre la regione ❸ sarà lossy. In ogni caso, dato che la regione ❸ contiene le stesse informazioni di allineamento temporale, il segnale audio ricostruito sarà virtualmente lossless. La frequenza di codifica di un file MQA è sempre basata su quella della traccia originale; quindi, per master 88,2 - 176,4 - 352,8 kHz tale frequenza è di 44,1 kHz, mentre per master 96 - 192 - 384 kHz tale frequenza di codifica è di 48 kHz. Con l'origami MQA tutti gli apparati di registrazione, produzione e distribuzione saranno gli stessi del normale CD. Inoltre, su un 1 TB si memorizzeranno 2.000 album 192 kHz / 24 bit MQA, mentre con il formato .flac gli album memorizzabili sono mediamente 640.

Vi sono due tipi di file MQA: MQA normale e MQA Studio. La differenza tra i due è solo relativa alla provenienza del file master.

Un file MQA normale è ottenuto da un master audio esistente e sottoposto a Master Quality Authenticated; il decoder indica in VERDE la correttezza del file in riproduzione. Un file MQA Studio è ottenuto da una registrazione sottoposta a Master Quality Authenticated e approvata dall'artista o dal produttore o verificata dal titolare del copyright; il decoder indica in BLU la correttezza del file in riproduzione.

In sintesi, il formato .flac, pur sempre eccezionale per la rivoluzione introdotta ai tempi dei file .wav, considera i file audio come dei semplici file dati.

Il formato MQA, invece, considera i file di dati come file audio e pondera quindi il contenuto musicale degli stessi, permettendo di memorizzare in massimo 1/3 dello spazio gli stessi contenuti rispetto al formato .flac; allo stesso modo questo permette di occupare al massimo 1/3 della banda di rete nei servizi di streaming rispetto a quanto richiesto per trasmettere un file .flac pur conservando, se non incre-

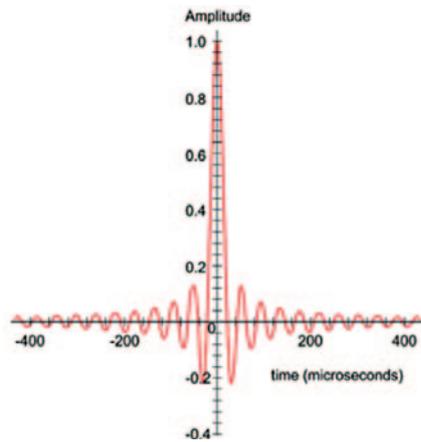


Figura 6: inerzia all'impulso audio

mentando come si vedrà in 2.2, la qualità audio di riproduzione.

Miglioramento della riproduzione audio con l'algoritmo MQA

In fase di riproduzione audio, come mostrato in Figura 6, le registrazioni digitali manifestano una certa inerzia agli impulsi; questo è evidenziato dalla riproduzione di un impulso e non dalla misura della risposta in frequenza.

Nella risposta all'impulso vi è tipicamente un pre-ringing (oscillazione), seguito dal picco dell'impulso, e poi segue un post-ringing. Questa pre e post oscillazione in cui il segnale manifesta una certa incertezza alla reazione è come una sfocatura. È opinione comune, tra i tecnici del suono, che la pre oscillazione contribuisca a rendere la riproduzione innaturale. Lavorando a frequenze di campionamento più elevate, tecnica utilizzata da alcuni costruttori, la pre e post oscillazione diminuiscono e potrebbe essere questo il motivo per cui molti avvertono una riproduzione più naturale, migliore. Una valida soluzione alternativa è quella di Sony che ha messo a punto il formato DSD e i SACD. L'implementazione, però, comporta l'adozione di una codifica ad 1 bit anziché in PCM, con la conseguente necessità di adeguamento di tutto l'equipaggiamento hardware dello studio.

In alternativa è possibile impiegare i filtri digitali che emulano quelli analogici (IIR), tipicamente non a fase lineare, i quali cancellano il pre-ringing e lo trasferiscono al post-ringing che di conseguenza incrementa fino a raddoppiare e più. È opinione accettata che l'assenza del pre-ringing anche al prezzo di un aumento del post-ringing conferisca al suono più naturalezza. Comunque, individuato il problema della sfocatura, sia che si lavori in alta risoluzione, sia che si operi con filtri IIR, con le tecniche tradizionali rimane sempre del ringing.

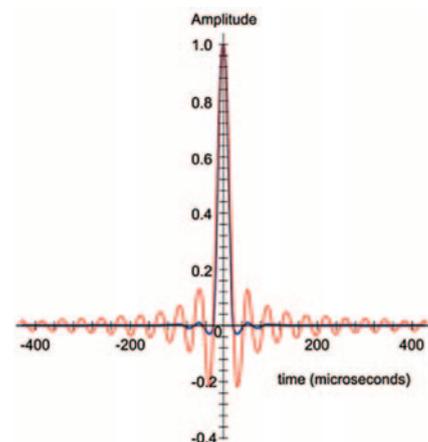


Figura 7: correzione del ringing con apodizzazione.

Gerzon, Graven e Stuart, con l'algoritmo MQA, hanno elaborato una tecnica di "messa a fuoco" (de-blurring in inglese) applicando al mondo audio tecniche di apodizzazione simili a quelle che si usano nei radiotelescopi, nei quali l'immagine non viene distorta perdendo l'informazione originale, ma viene per l'appunto solo messa a fuoco. In Figura 7 la curva blu riporta la risposta all'impulso dopo l'applicazione della apodizzazione secondo MQA. La curva rossa è invece la risposta all'impulso corretta in modo tradizionale. L'effetto dell'algoritmo MQA è evidente: il pre e post-ringing è ridotto di circa l'80%.

MQA, contrariamente al DSD dei SACD di Sony, non richiede un hardware dedicato in sala di registrazione, e si applica a qualsiasi registrazione PCM standard, anche già esistente.

L'algoritmo MQA di de-blurring è realizzato in due fasi: una avviene in studio di registrazione, ed una seconda avviene a livello di unità DAC. Qui vengono utilizzate le informazioni di Master Quality Authenticated introdotte in studio di registrazione.

Da qui MQA letteralmente, mai traduzione dall'inglese fu peggiore, Qualità Master Autenticata.

L'orecchio, il campionamento a 44,1 kHz, a 192 kHz, e 24 bit di risoluzione

Qualche cenno sui numeri utilizzati in precedenza che ricorrono quando si parla di "risoluzione CD" o di "alta risoluzione" in campo audio.

Il punto di partenza è lo studio dell'orecchio umano.

La banda audio percepita dall'orecchio umano è 20 Hz - 20 kHz

Semplificando al massimo, all'interno dell'orecchio si trovano: padiglione,

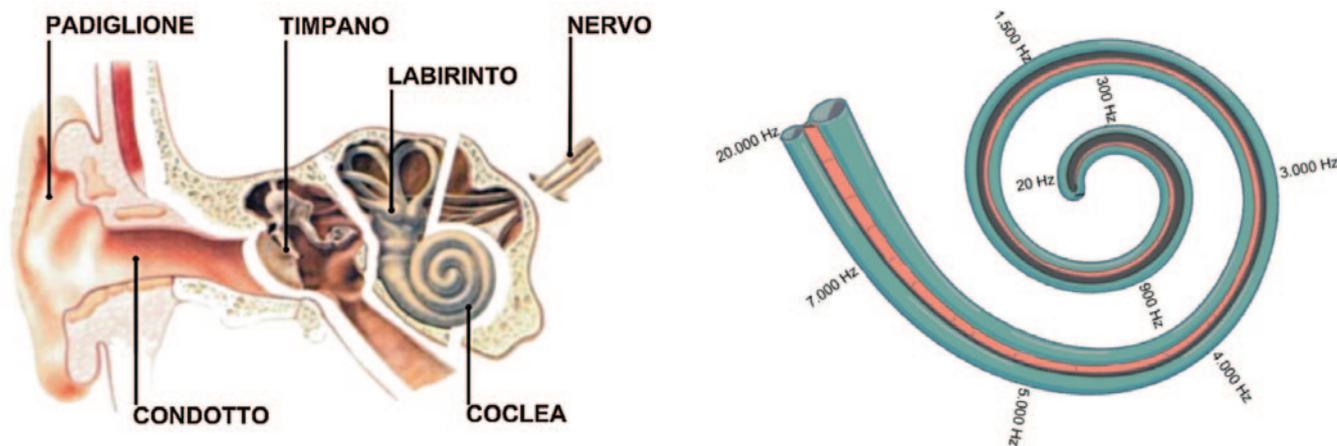


Figura 8: la coclea, schematizzata nella figura di sinistra, è una sorta di tromboncino che si chiude a chiochiola.

condotto, timpano, labirinto, coclea. Quando si nasce, nella parte terminale della coclea i recettori catturano i 20 Hz e nella parte iniziale della stessa, dove questa è più aperta, i recettori catturano i 20 kHz. Per gli over 50, che hanno capacità uditive ridotte rispetto alla nascita, la banda si riduce orientativamente a 20 Hz - 10 kHz; va detto però che non si traduce in una perdita di metà dello spettro uditivo perché l'orecchio lavora in modo logaritmico, ossia in ottave e, dividendo in ottave i 20 kHz, si hanno le ottave 20.000 - 10.000 - 5.000 - 2.500 - 1.250 - 625 - 312,5 - 156,2 - 78,1 - 39 - 19,5 Hz. Ogni trattino è una ottava, in totale 10 ottave.

Perciò, superati i 50 anni, si perde solo la prima ottava di sensibilità sulle 10 che si hanno alla nascita e non metà della capacità uditiva, continuando a sentire perciò 9 ottave; in più, secondo gli studi di alcuni ricercatori, si continuano comunque a recepire i rumori impulsivi trattati al punto 3.2. Il comportamento dell'orecchio appena descritto serve all'essere umano per la comprensione; l'orecchio è uno dei 5 sensi. Per approfondimenti: bit.ly/3mdWwJ1

La risoluzione temporale dell'orecchio

Vi è un'altra importantissima caratteristica dell'orecchio, ignorata, sconosciuta ai più. L'orecchio, quando usato come organo di sopravvivenza e non più di comprensione, esibisce una sensibilità che nulla ha a che vedere con la banda 20 Hz - 20 kHz.

È la sensibilità ai rumori impulsivi, in inglese *twig snapping sounds*. Si tratta, per esempio, dello scricchiolio secco e impulsivo che si percepisce quando si spezza un ramoscello, oppure dello scoppiettio della legna secca che arde sul fuoco; si tratta di una serie di *scrick, stick, steck, stock*. A nessuno viene in mente di chiamarli suoni, vengono chiamati rumori: "il rumore della legna sul fuoco" e non il "suono della legna sul fuoco".

L'orecchio riesce a distinguere rumori impulsivi che distano tra loro 6 microsecondi (ms).

6 ms equivalgono ad una frequenza di campionamento di (tempo = 1/frequenza) 181 kHz e l'insieme apparato uditivo - cervello li recepisce e li discrimina.

Approfondimenti:

<https://bit.ly/39iO3AT>

<https://bit.ly/3l9m48P>

Il range dinamico dell'orecchio

Il range dinamico percepito dall'orecchio, ossia la differenza tra il suono più flebile e il suono più forte, è dell'ordine di 96 dB ma, per brevi periodi, può sopportare anche 120 dB; ma se si prolunga il tempo di esposizione o si sale l'orecchio si danneggia.

Riassumendo, dallo studio della anatomia dell'orecchio è noto che:

- si percepisce una banda 20 Hz - 20 kHz;
- si distinguono rumori impulsivi che distano fra loro fino a 6 ms;
- il range dinamico è di 120 dB.

Nyquist e Shannon, un fisico ed un matematico, hanno dimostrato che digitalizzando un segnale di qualsiasi natura sia esso video, sonoro, ultrasuoni, UHF, VHF, solo se campionato ad una frequenza almeno doppia della larghezza di banda del segnale stesso non si ha perdita di informazione, nel senso che lo si può ricostruire perfettamente. Pertanto, facendo riferimento al punto 3.1, essendo la banda utile 20 Hz - 20 kHz, basterebbe campionare a 44,1 kHz per conservare la comprensione (poco sopra al doppio per altre questioni che esulano da questo articolo).

Però, facendo riferimento al punto 3.2, l'orecchio cattura impulsi separati tra loro da almeno 6 μs, perciò se non si vogliono perdere queste informazioni si deve campionare ad almeno 181 kHz (1/181 kHz = 6 μs).

Dato che 192 kHz è il multiplo di 44,1 kHz più vicino, ne risulta il numero familiare di 192 kHz di frequenza di campionamento (Tabella 2). Dato che l'orecchio ha un range dinamico di 120 dB, 20 bit sono sufficienti per coprire l'intervallo dal suono più flebile al suono più forte; i computer però usano i bit a multipli di 8, quindi 8, 16, 24 ed ecco l'altro numero familiare di 24 bit.

Pertanto:

- a) per non perdere la comprensione si campiona ad una frequenza di 44,1 kHz, e questa è la "risoluzione CD";
- b) per non perdere i rumori impulsivi si deve campionare ad una frequenza di 192 kHz, e questa è detta "alta risoluzione";

Tabella 2 - L'elaborazione digitale del segnale lavora con i bit

1 bit = 2 livelli	con 1 bit si hanno 2 livelli, solo 0 o 1	6 dB di range dinamico
8 bit = 256 livelli	con 8 bit si hanno 256 livelli, da 0000000 a 11111111	48 dB di range dinamico
16 bit = 65.536 livelli	da 0000 0000 0000 0000 a 1111 1111 1111 1111	96 dB di range dinamico
24 bit = 16.777.216 livelli	non c'è spazio ma dovrebbe essere chiaro lo sviluppo ...	144 dB di range dinamico

Nel passaggio da livelli a dB si è utilizzata la formula $20 \cdot \lg(\text{livelli}) = \text{dB}$; quindi, per esempio, 2 livelli equivalgono a 6 dB ($20 \cdot \lg 2 = 6 \text{ dB}$).

c) per rilevare la gamma dinamica dell'orecchio servono 16 bit (96 dB), quella della "risoluzione CD", mentre per rilevare l'"alta risoluzione" servono 24 bit (144 dB).

Quindi 44,1 kHz / 16 bit per la "risoluzione CD" e 192 kHz / 24 bit per l'"alta risoluzione".

Il formato .wav ed il formato .flac e le informazioni nelle tracce digitalizzate

Il formato .wav colleziona tutti i campioni della digitalizzazione senza nessuna post-elaborazione. Il formato .flac ha reso possibile la compressione del formato .wav senza alcuna perdita di informazione.

Una traccia audio campionata a 192 kHz / 24 bit e non compressa restituisce una banda audio 20 Hz - 96 kHz senza perdita di informazione.

Semplificando al massimo, si ricorda che l'uomo usa per la comprensione la banda 20 Hz - 20 kHz, per i rumori impulsivi la banda 48 kHz - 96 kHz e, per entrambe, la banda 20 kHz - 48 kHz. Fino a 20 kHz si hanno suoni, dopo ultrasuoni, così detti perché non udibili ma percepibili in quanto comunque utilizzati dal sistema orecchio-cervello durante l'ascolto.

Si noti che nella banda di frequenze 20 Hz-96 kHz si hanno 12,2 ottave:

- da 20 Hz a 24 kHz ci sono 10,2 ottave;
- da 24 a 48 kHz si ha 1 ottava;
- da 48 kHz a 96 kHz si ha 1 ottava.

Per inciso, anche se potrà apparire non intuitivo, la banda 48 kHz - 96 kHz non corrisponde a 1/2 banda ma solo a 1/12 perché l'orecchio lavora in modo logaritmico, ossia in ottave.

L'uomo comprende utilizzando le prime 11 ottave e colloca temporalmente usando le ultime 2 ottave; pertanto all'interno della banda di frequenze 20 Hz - 96 kHz usata per l'audio, non tutte frequenze hanno lo stesso peso e questo è il presupposto fondamentale alla base di MQA.

Come si implementa MQA

Per i costruttori di unità digitali, la funzione MQA è facilmente integrabile ed è basata su tecnologia XMOS. L'implementazione non prevede che il progetto, l'architettura, dell'unità DAC siano condivisi con MQA Ltd. Si tratta di un ricevitore USB che implementa MQA; l'unità DAC non deve fare nulla.

Come suona MQA?

Nelle valutazioni relative alle tecni-

che digitali applicate, e pertanto anche quando si valutano i formati di memorizzazione, è importante fissare l'ambito ossia le condizioni al contorno.

Descrivendo l'origami MQA e il filtro di de-blurring - apodizzazione - messa a fuoco MQA si sono evidenziate le caratteristiche degli algoritmi e i comportamenti a livello matematico.

Di fatto l'esame è limitato ai chip che offrono le varie funzioni e non ai dispositivi che vengono realizzati con tali chip. Nel caso dei DAC la disamina finora ha riguardato i chip di base ma non le unità DAC nella loro interezza. In un'unità DAC, dove si trovano alimentazioni, circuiti di ricezione, stadi di conversione corrente tensione, filtri, il chip contribuisce normalmente solo per un 10% al risultato finale.

Inoltre, il suono usufruibile dall'utente finale è prodotto dall'intera catena di riproduzione dove la sorgente è seguita da un preamplificatore, da un amplificatore di potenza e termina con dei diffusori, il tutto interconnesso da cavi e alimentato alla tensione di rete.

Il fine ultimo è quello di ascoltare in cuffia o in una stanza, pertanto la valutazione "come suona MQA?" deve essere considerata sulla intera catena audio e non sul singolo dispositivo o sul singolo chip. Il confronto inoltre deve essere alla pari:

- stesso brano;
- stesso hardware;
- diverse risoluzioni.

Il sito bit.ly/3qdsCY7 offre tracce audio gratuite in normale/alta/altissima risoluzione dello stesso brano e in formati diversi, .flac, .wav, .dsf, .dxd, .mqa.

Deve essere utilizzato lo stesso hardware, ossia un computer preferibilmente con assetto audio (attenzione all'alimentatore), Jriver, Audirvana o Foobar come player, uscita USB e cavo USB a basso jitter, un'unità DAC con basso jitter, basso rumore di fondo generale, assenza di ripple in cassa. Solo allora ha senso il confronto tra i vari formati.

È anche importante impostare lo stesso livello di ascolto; non è insolito infatti avere lo stesso brano con livelli diversi passando da una risoluzione e/o un formato all'altro. Normalmente si pensa che una traccia con livello più alto abbia più dinamica, oppure che sia equalizzata nella parte bassa e/o nella parte alta. Non è così; si ricordi quindi che prima di fare i confronti è necessario allineare i livelli.

Non va utilizzato Bluetooth per collegarsi allo streamer o al player poiché viene tagliata la banda audio per le caratteristiche tecniche di questa tecnologia.

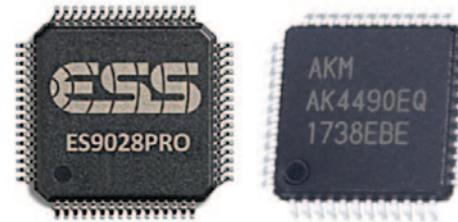


Figura 9: due noti chip utilizzati come DAC di base.

Sistemata la catena per le prove, tra l'algoritmo MQA e la compressione FLAC dovrebbe essere abbastanza facile riscontrare che non vi sono differenze; il contenuto audio infatti è lo stesso, così come il ritmo temporale.

Sul de-blurring MQA invece dovrebbe essere riscontrabile una maggiore naturalezza dei brani musicali e una maggiore messa a fuoco e collocazione 3D dell'immagine sonora, in quanto sia pre che post-ringing sono ridotti dell'80% rispetto agli altri formati.

Per valutare la messa a fuoco sono preferibili brani musicali con viole, violini, violoncelli, dato che l'orecchio è molto sensibile alle variazioni di timbro e ritmo proprio lì dove suonano gli archi. Tanto più naturali e distinti saranno gli archi e tanto più evidente sarà la messa a fuoco temporale. Ovviamente l'impianto utilizzato nella prova deve essere messo a punto, nel senso che è inutile cercare di percepire la messa a fuoco, il giusto timbro, il giusto ritmo, se il tasso di jitter presente è elevato, se vi è ripple residuo, se vi sono loop di massa, se sono catturate interferenze a causa di scarse schermature, ove vi sono trasformatori di alimentazione che irradiano, peggio se del tipo switching.

Conclusioni

Si è visto che se non si vogliono perdere i rumori impulsivi si deve campionare ad almeno 192 kHz, ossia in Alta Risoluzione. Pertanto l'alta risoluzione è certamente necessaria per un ascolto fedele. Nulla c'entra la banda 20 Hz - 20 kHz che comunque resta fondamentale per la comprensione.

Si è visto poi come un ascolto con pre e post-ringing ridottissimo conferisce maggiore naturalezza alla riproduzione.

MQA affronta l'alta risoluzione ponderando i contenuti alle varie frequenze. I contenuti sono preservati per la riproduzione audio e al tempo stesso la dimensione dei file audio è ridotta almeno di 1/3 rispetto ad altre soluzioni lossless.

Mario Canever