

J. Robert Stuart - MQA: Questions and Answers

- Stereophile 誌記事より抜粋

数ヶ月前に MQA に関する様々な質問に対して Bob Stuart がひとつひとつ丁寧に答える長文の記事が Computer Audiophile にアップされているのを見て大いに感心させられたのですが、さらに追い討ちをかけるように出て来たのがこの Stereophile 誌の Q&A 記事です。概念的な話は以前訳した Absolute Sound のインタビュー記事が解りやすいと思いますが、Bob がこうして本腰を入れた技術解説に取り組むのは業界認識が一様とは行かない現状の証左なのかも知れないと、喜んでいられない気分です。僕など、MQA 技術の構想を聞いて音を確認すれば納得して終わるのですが、どうもオーディオマニアの世界では疑心暗鬼や異論が蠢くようで、例えば「本当にロスレスなのか詳しい説明を」と追求姿勢です。それは「デジタル伝送部だけのロスレスにこだわって何の意味があるの」というもっと大きな展望が MQA としてのテーマであることを認識すれば、open-mind な意思疎通はできることなのですが。この Q&A 記事を丸ごと訳出するのは量的にも荷が重いので、骨となる部分だけでも拾い出してまとめてみたら、もしかして役に立つこともあるのだろうかと考えて翻訳を試みることにしました。まとまりとして期待通りかどうかはともかく、参考になるところが少しでもあればという思いです。

2016/9/09 伏木 雅昭

#

Temporal blur のこと

我々が時間領域の滲み(blur)のことを議論する時、チャンネル内あるいはチャンネル間のタイムベース量子化のことを話しているのではありません。

時間領域の滲みについては測定標準がありませんが、我々の用語定義については明確で理解しやすいものと考えます。普通の伝送システムではフィルターもしくは減衰動作による分散特性があります。End-to-end のインパルス応答が信号もしくは受信機（聴いている人）の感度に対して不十分だと、時間波形の細部が滲んで曖昧になります。光学の世界での滲みもまさに同様に、レンズ設計、媒体内での光の拡散、画像処理に関係します。電子分野では、このことはオシロスコープ設計者にはよく理解されています。

人間の聴覚が周波数よりも時間に敏感らしいとの注目に値する証拠が神経科学の研究から出て来ており、人間の聴覚はフーリエの時間・周波数不確定性より能力が高く、また時間的微細構造に対する感度は同一帯域を扱うリニアシステムよりも繊細であると我々はそれを理解します。音について人間の聴覚に重要な細かさは時間軸上 $5\mu\text{S}$ ほどの短さであると思われます。ここで何より重要なのは、こうした時間上の微細な事象が必ずしも高周波成分から生まれるのではないと認識しておくことです。音は様々な対象物からマイクに到着し、そこには残響もあり、声や楽器が点音源ではないと認識します。面白いことに、このレベルの感度というのは人間の聴覚限とされる $\sim 18\text{kHz}$ とは相関していません。

リニアなアナログシステムの場合、カスケード要因が高域ロールオフをもたらすので、中心点（群遅延）を動かして分散するあるいは細かく分離された事象を融合できる関数によって時間領域の細部は平滑化されることが理解できます。図 4 参照。

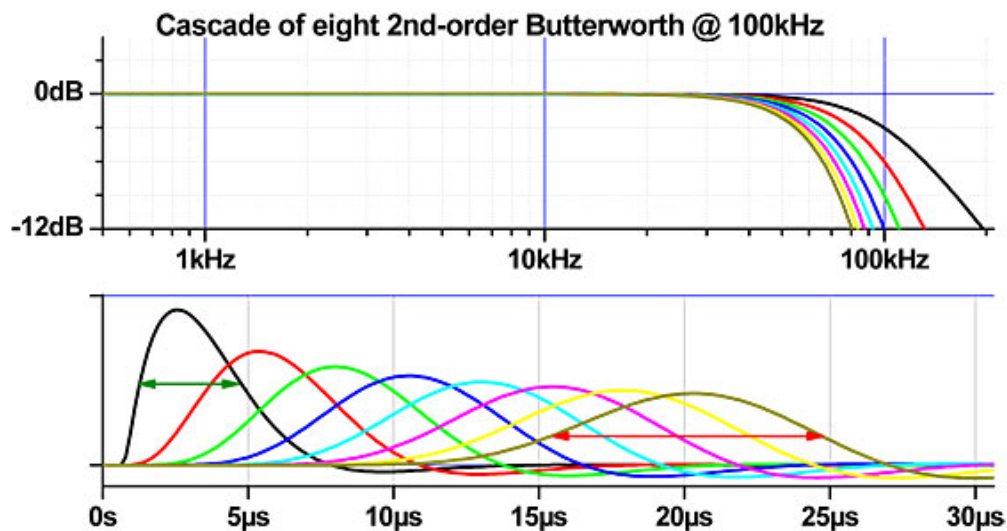


図 4: LPF カスケードの周波数及びインパルス応答

録音系トータルを考えた場合、考えられる構成部分それぞれの設計者は 100kHz までの帯域をカバーすべきであるということかも知れませんが、それは現実的ではなさそうです。ごく最近まで、各構成部が 30kHz くらいまで特性がフラットなら妥当と見なされていました。しかしそれは演奏者からリスナーまでの全旅程がマイク、プリアンプ、ミキサー、コンバーター、プリ及びポストフィルタ、再生のプリアンプ、パワーアンプ、そしてトランスデューサーのカスケードになっているとなると、その数字は時間領域の観点からはとても妥当とは言えないものとなります。こうした系では $8\mu\text{s}$ のバジェットはすでに使い果たしていると判りますし、ひとつの部分を 30kHz に制限するだけでも、外挿 extrapolation によって（[文献 2](#) の図 3）、全バジェットが消費されると判ります。

この議論は信号の時間領域での滲みに基づくものであって、リスナーの人間が 30-100kHz 帯域の高調波から恩恵を受けるという（ありそうにない）要件に対するものではないと認識することが最重要です。システムを広帯域にすべきであるという論拠は高域成分がその理由なのではなく、**我々の神経が 20kHz 以上をフィルタすることとで発生するオーディオ帯域内の変化に μs レベルの感度を持っているからなのです。**最終的なことは、これはアナログ領域で定義されるもので、音は空気の中のアナログなものであって、デジタル保存チャンネルが介在するならそれはこのフレームワークに適合する必要があるということです。

標本化システムと Temporal blur

標本化による符号化システムのルーツを遡ると、シャノンのエレガントな再構築定理やナイキストのチャンネル容量定理に出会います。標本化された信号は $f_s/2$ 以上の周波数成分がなければ、明快に再構築できます。それは $T=1/f_s$ 秒ずつ離れた一連の点毎にその信号値を取り込めば完全に決定できることで、 $f_s/2$ での完璧な LPF を通せば再構築可能です。この概念構造は人間のリスナーにとってはいくつも問題があります。まず、厳密に数学的で完璧な LPF（いわゆる brick-wall）はあり得ません。相互作用的にそれが存在したとすれば、信号は無限の時間を使って変化することになりますが、我々の関心は時間的にもっと急速な信号変化にあります。厳格な技術工学かあるいは怠惰な思考によってか、我々は「 $f_s/2$ 以上の周波数を含まない」要件に対して「汝、ブリックウォールを持つべし」という答を持ち込んでしまったということです。

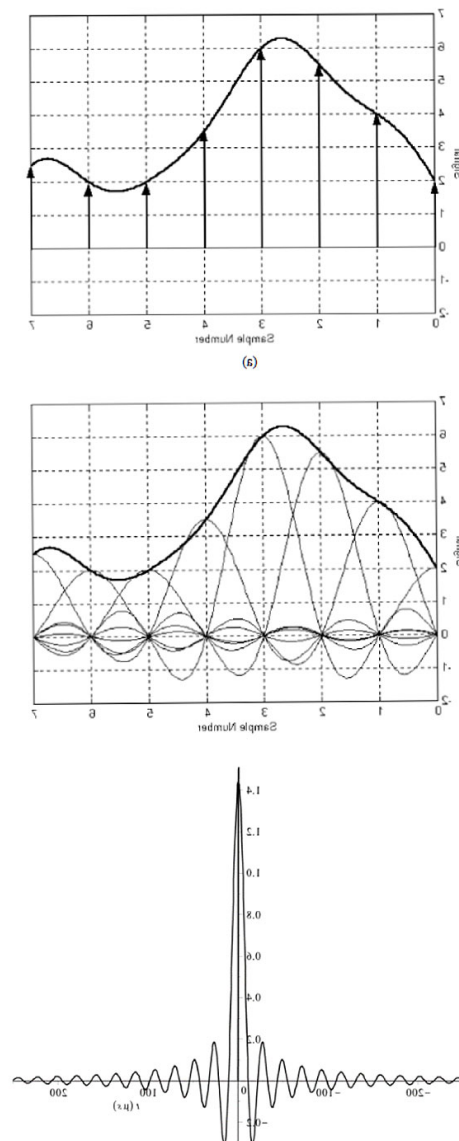


図 5: 標本化システム

ブリックウォールのフーリエ変換は図 5 のようによく知られた sinc 関数($(\sin x)/x$)の時間応答を示します。この関数には利点・欠点があり、

- 有利な面としては、センター点を除けば各サンプル点での sinc 値が 0 となること
- 不利な面としては、sinc フィルターによりシステムの時間応答が $\pm\infty$ から滲むこと（物理世界には特性がこれと合致するものではありません。）
- リスナーの人がリングングの周波数を聞くことはないかも知れませんが、エンベロープ全体に対しては敏感であるという注目すべき証拠があります。
- この関数の不確かな恩恵としては、フィルターの位相特性がリニアで、80 年代中頃以来 A/D や D/A 変換 IC では近似的 sinc フィルターとして便宜性があり、一般化していることです。

しかし正真のデジタルオーディオシステムとしてはブリックウォールは実在しないので、狭義の観点では決して理想的ではありません。では、A/D や D/A のフィルターが理想でないとな何が起ころのでしょうか？もし元信号に $f_s/2$ 以上の周波数成分が含まれていた場合、下方帯域にエリ阿斯が発生します。エリ阿斯の存在そのものは悪ではありません – 正規の信号が間違った周波数で再現されているということなので – しかしサンプルレートが 44.1kHz のような低い A/D 変換の場合となると、22kHz の上あたりに重大なエネルギーが入って来る可能性があり、超音域成分が擬似的に可聴帯域に発生することは困るので、最も厳しい制約を抱えることになる訳です。にもかかわらず、A/D 変換器は急峻な $1/2$ 帯域フィルターを使用しており、そのため **18~22kHz 帯域に下方エリ阿斯を許容している状況です。**

この下方エリ阿斯を回避するひとつの方法はサンプルレートを上げることで、それによりコンテンツがナイキストを超えないようにするか、フィルターの傾斜を弱める妥協を許すことです。

信号をデジタルで取り込んでしまっている以上、その制約も抱えたままです。シャノン定理では単位時間当たり $1/f_s$ だけの自由度があることを思い出してください。但し信号連続性が途切れると、その有効性が失われます。デジタル信号処理で生成される信号には再量子化、圧縮、クリッピングなど、帯域制限されていないものもあり、結果的に我々の現行再生システムはデジタル及びアナログ領域で全くロスレスとはなっていないのです。

デジタルをアナログに変換し直すところでは再構築を使用します。サンプリングの伝統的な認識では、周波数特性をフラットに保ち、情報エリ阿斯を防ぎ、あいまいさを残さないためにはブリックウォールフィルターが必要としてきました。もっと最近では、特に Craven 論文以降、例えばアポダイジング特性のような、より緩やかな再構築フィルターを用いたアップサンプルが有効と理解されています。

今では高性能 D/A 変換器はフィルターを選ぶ機能を提供していて、「測定対試聴」のような記載なども見られます。こうした変換器を作ろうという起動力に一役買っているのが、A/D からリングングを意識的に除去して、但しそれ以降の副産物 artifact は最終結果のアナログに預けようという動向です。追求した妥協策には高域特性をだら下がりさせることや上方エリ阿斯もあります。しかし冪等でない恣意的なフィルターを使用すれば、アナログ出力が元信号の再構築にならず、従ってその恣意性任せに必ず陥ることになります。その出力は最早入力 of 忠実な再現とは言えず、典型的には高周波特性に問題を抱えます。

MQA でひとつ重要なポイントは、サンプリング及び再構築フィルターがエンコーダーとデコーダーで相補関係になっており、従って最終アナログ出力がスタジオでモニターした音の忠実なコピーとなることを確実にしている点です。

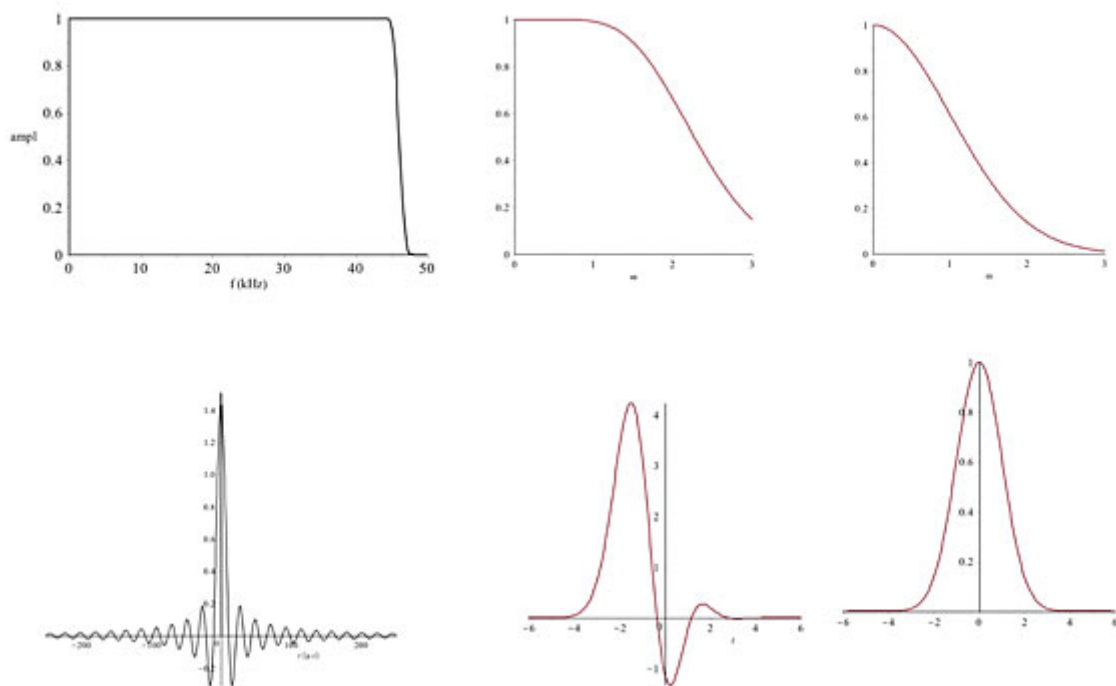


図 6: カーネルの異なる標本化システムの end-to-end 特性例。

上のグループは特定チャンネルの周波数特性で、下はそれぞれのインパルス応答を示す。左は sinc 型直線位相チャンネルの典型。右側は Gaussian フィルターを両端に持つチャンネルで、インパルス応答は理想に近いが、周波数特性は早々に垂れ下がっている。Gaussian フィルターは画像伝送やオシロスコープのように波形が問題となる用途では理想型だ。人間の聴覚機能はまた異なる時間・周波数バランスを持っており、中央の図のような中間的なカーネルを用いることで音を最適化して伝送できる。

音楽のスペクトラム

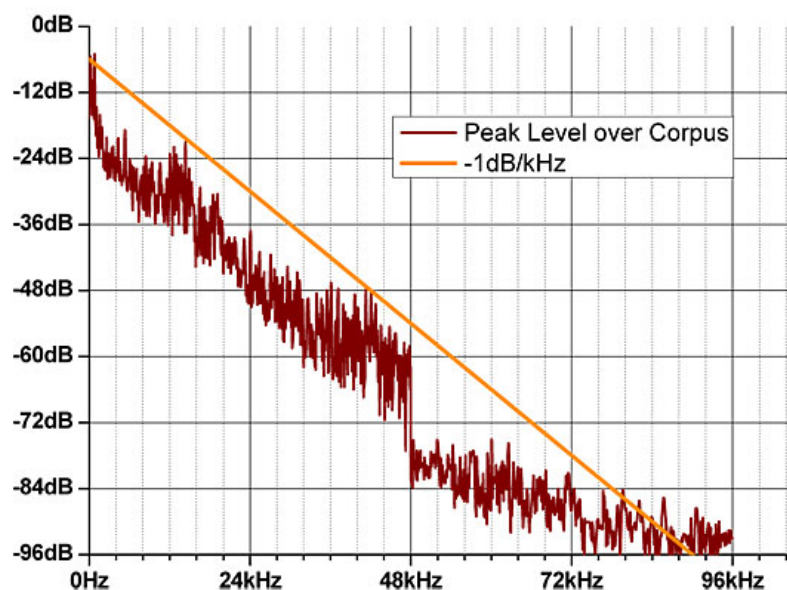


図 14: 96k 及び 192kHz 録音の集積でのピークスペクトラムレベル

ハイレート録音分析をまとめた図 14 で判るように、多くの種類の音楽で 20kHz 以上に重要な成分があります。楽器の特性でひとつ注目すべき共通点は周波数上昇につれてそのパワーが落ちていくことで、時に大きく減衰します。楽器が 20kHz 以上の音を発するとしても、トランスペアレントなシステムが必ずしもそれを再生できなければならないということにはなりません。大切なのは帯域を絞るために用いた手法がリスナーに検知できるか否かなのです。

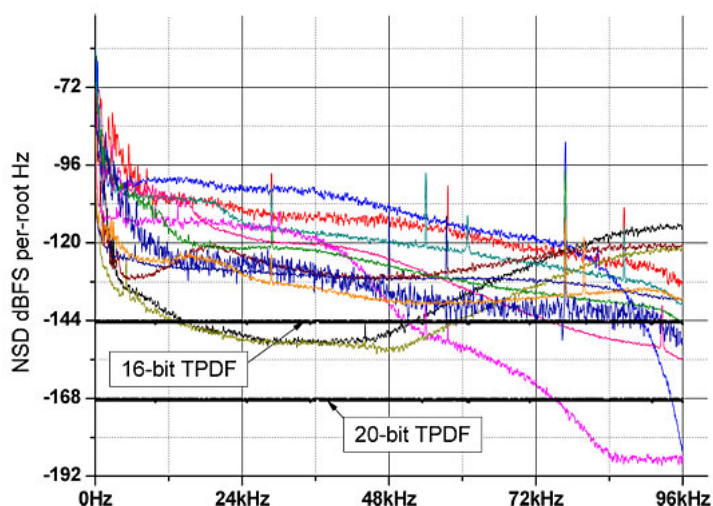


図 15: 192kHz/24bit 市販ソフトでの背景雑音の例。192kHz/16bit 及び 20bit 量子化に対する TPDF ディザイナ雑音も示す。曲線は 1Hz バンド幅での雑音-スペクトラム密度のプロット。

録音での背景雑音測定を見ているのが図 15 で、60 年前のアナログテープの未加工再発物や最新デジタル録音まで含んでいます。当然ながらマイクや演奏会場の雑音もそこには組み込まれ、アナログテープノイズを含むものもあります。最良のテープレコーダーのノイズフロアでも 16bit 理想チャンネルのノイズフロアより上になります。ここで 20bitPCM チャンネルはこれら録音を取り込むには余裕で妥当なもの、従って 32bit 精度から明確な恩恵というのは見つからないと認識しておく価値があります。ひとつひとつのどの録音を見ても、ピークとノイズの曲線が三角形を形作っていることが判ります。三角形の形状（鋭角、鈍角など）は特定の素材により異なります。それでもこの洞察により、我々はどんな実際の録音にも以下の要素があると考えることができます。

- 音楽演奏の前後にも持続している背景雑音
- 限られた割合で革新を取り込む（ほぼスルーレート限界）ピークレベル
- それより高い帯域では信号の変動部分が雑音レベルより下になる周波数
- シャノン図式でプロットして、音楽中の情報分がそのまま判る領域

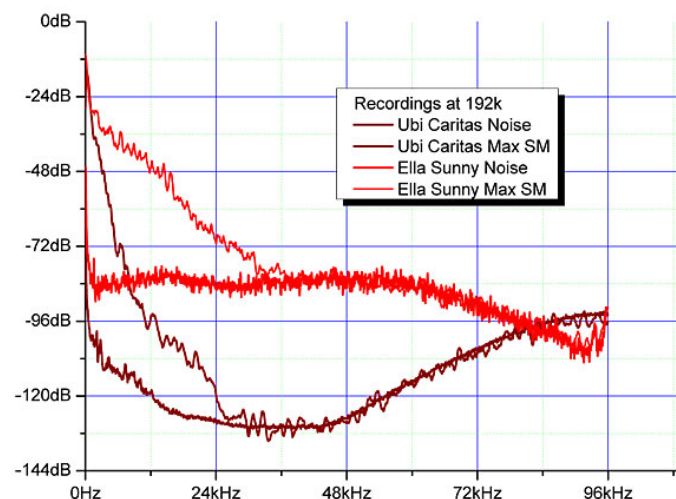


図 16: 2 種類の録音でのピークと雑音のプロット

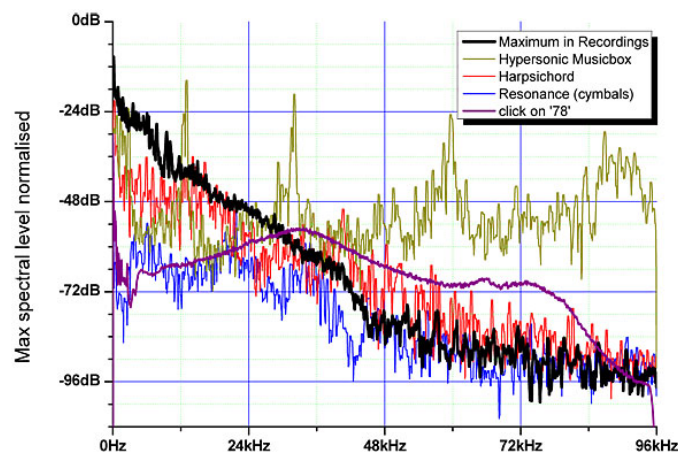


図 17: 192kHz 市販ソフトの大量調査でのピークレベル

図 17 の黒い曲線は市販 192kHz の音楽ファイルをかなりの量調査して、ピークスペクトラムのピーク値を示したものです。全般として 192kHz でリリースされた録音は 96kHz でマスターされたものよりレベルが低めです。その詳細として青い曲線を見ていただくと、ミックス口による優れた広帯域録音でのシンバルの衝撃音のピークスペクトラムです。他の 3 点はやり過ぎなくらい徹底した録音（非売品）のピークスペクトラムで、我々が実験に使用したものです。

- 紫色は 78 回転シュラック盤のサーフェスノイズ
- オリーブ色は仁科博士から提供された極超音域ミュージックボックスを特別にトランスファーしたもので、そのピークレベルを示します。
- 赤色はサンケン 100kHz マイクの耐久試験としてハープシコードの蓋板下の爪に近いエリアにマイクを置いた近接録音でのピークレベルです。

Lossless 性に対する疑問

「MQA は本当にロスレスですか」という質問はロスレスが最善のものという想定に立っていることが多いのですが、ロスレスがやっていることは単にいくつかのビットを取り出してそれらを時間と場所の異なるところで再生するということです。皆さんが望むのがそれだけなら、FLAC で充分でしょうし、MQA の必要はないでしょう。MQA のチームはロスレス圧縮だけでなく、ロスレスプロセスやデータ埋め込みについても理解をしています。(MQA チームが初めてロスレス圧縮システムをインプリして RIAA にデモを行ったのは FLAC が生まれるより 10 年前のことです。)

さらにアナログからアナログへのシステム経路は他の技術要素も加わってもっと精緻です。ロスレスが扱うのはデジタル領域内のデータで、**我々の経験では最大の問題はデータをアナログから取り込むところと可聴な損傷を最小に抑えてそれをアナログに戻すところです。**この俯瞰図を理解しないと、MQA は奇妙なものに見えるでしょう。

MQA が対処しようとしている問題は、アナログ信号を如何にして異なる時間と場所に移動させるかです。そのアナログ信号はミキシング卓で制作者が聴いたものであり、あなたが自分のスピーカーから再生したいと思っているものです。

多くの録音及びマスタリング技術者が MQA は従来の方法に著しい改善をもたらし、彼らがオリジナルセッション時に現実に聞いたあるいはそう記憶している音を、またアーカイブ素材の場合はアナログテープレコーダーの音を再現していると証言しています。

音の「折り紙」理論について

MQA の構造には柔軟性がある、「16bit のオーディオデータはベースバンドへ」との表現で含意するモデルには正確には相当しません。24bit MQA ファイルは少なくとも 3 通りのオーディションが出来ます。

- 16bit にトランケートしてデコーダーなしでオーディションする
- 16bit にトランケートしてデコードする
- 24bit から完全デコードする

回復された 0~20kHz のベースバンド信号はこれら 3 通りの提示間で異なることを知っておく必要があります。最大限広義には、上位 16bit は 0~20kHz 帯域のほとんどあるいはすべてを伝送し、残りのビットでベースバンドの解像度及び／あるいは周波数帯域を改善します。但し、これらのビットが解像度強化とか帯域拡大を提供するというものではありません。MQA というのはそうしたことよりもっと精巧です。

性能論議をする際、3 通りの提示方法についてどのケースの話なのか明確にしておくことが大切で、3 番目の場合なら 24bit 完全デコードですとなります。

ここから「折り紙」のおさらいです。

「Music Origami」はサンプルレートの高い信号を $f_s=44.1/48\text{kHz}$ に低減する MQA の処理を日本の紙を折る芸術になぞらえた呼称です。図 21 から 23 は 192kHz 音源に対する処理動作を示します。ここでは処理が 2 次元で描かれていますが、実際には 3 次元構造です。

第 1 の折り畳み（図 21）では転送レートを 192kHz から 96kHz に低減し、第 2（図 22）で 96kHz から 48kHz に折り畳みます。折り畳み処理はフィルターではないので、先天的なサンプルレートとビット長は維持されます。トランスポート中は折り畳まれているサンプルがより多くの情報を運びます。この第 1 の折り畳みは MQA 語彙としては「カプセル化」と言って、理想的には P 点より高い周波数で行います。そうでない場合には別の対処法と選択（後述）があります。

プロセスとしては階層的にスケラブルで、例えば音源が 352.8kHz ファイルなら 44.1kHz に落とすまで 3 回折り畳み処理します。同様に 96kHz 音源なら、図 22 のロスレス処理から始めます。MQA では折り畳みのタイプを 1 オクターブ高くして倍速転送を実現する選択ができる点でも階層的にスケラブルになっています。

第 2 の折り畳みでは、図に描かれたようにオリジナルのアナログ信号（青線）のノイズフロアに対して 24bit 符号化での有効ダイナミックレンジの多くをランダムノイズで占めていることから、ロスレス処理が可能です。そこで第 2 の処理は 24kHz 以上の信号情報を 24kHz 以下のノイズフロア内に可逆的にデータ埋め込みする技術を使います。

原理的に環境雑音やマイク自身の雑音（加えてアナログマスターのテープノイズ）が積み重なるため、アーカイブ録音で 16bit ダイナミックレンジを上回るのは言うまでもなく僅かです。さらに、ノイズの中の信号でもノイズレベル（図 21 のオリーブ色曲線）の 10dB 下までなら聞こえるという事実と従って 19~24bit のところには何ら有効情報がないということがそれに加わる訳です。

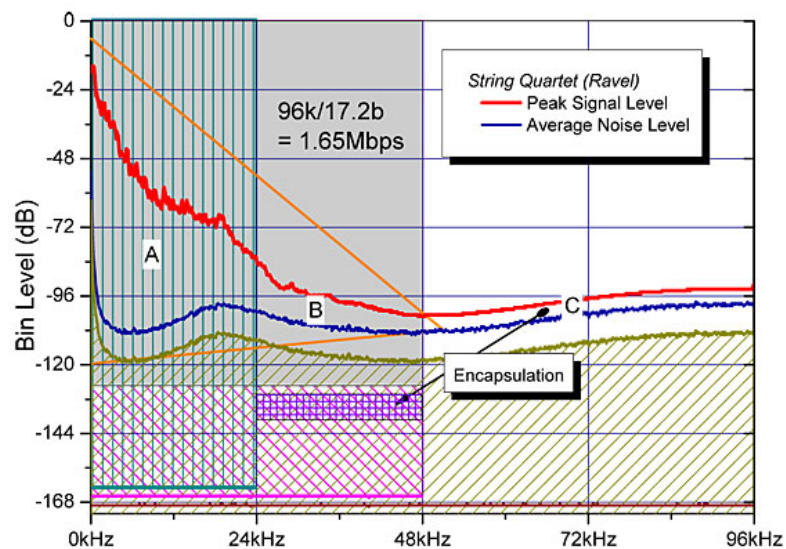


図 21: 第 1 の折り畳みは E（カプセル化）。
転送レートを半分にし、C の信号成分を B エリアの雑音レベルの下に保存する。

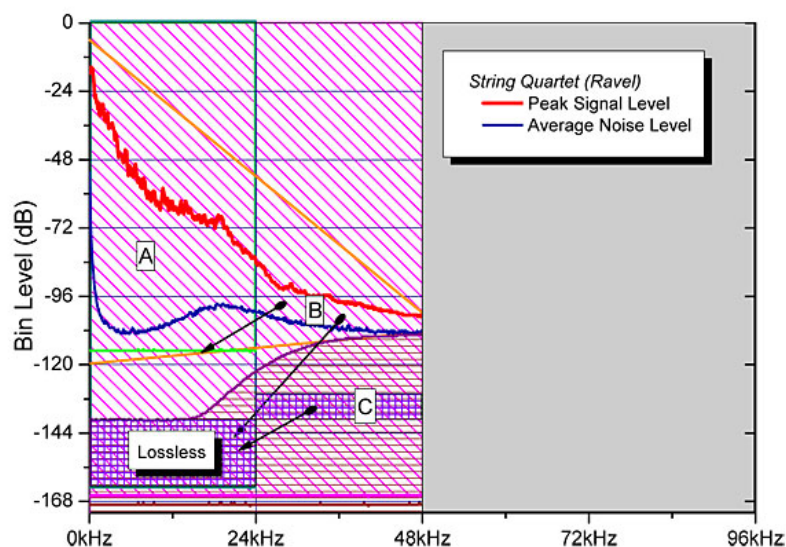


図 22: 第 2 の折り畳みは L タイプ。
転送レートを更に半分にし、B エリアの信号を C と共に
ベースバンドスペクトラムの雑音の下にロスレスで埋め込み、隠す。

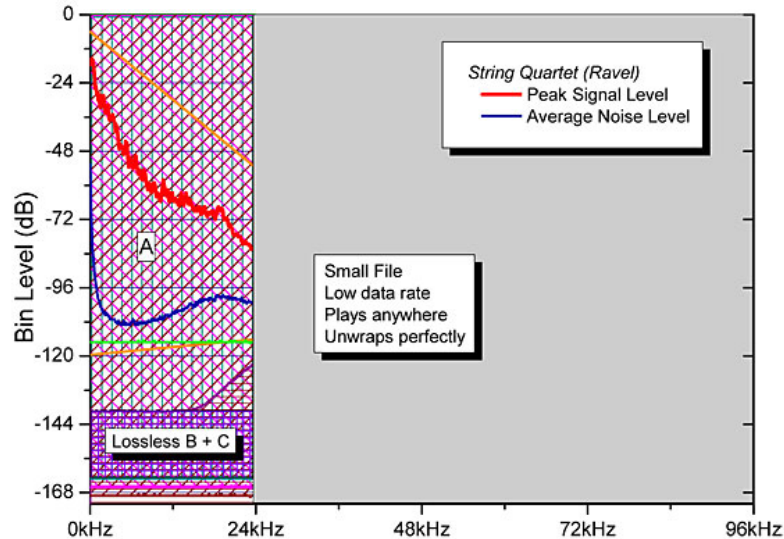


図 23: MQA ダウンロードファイルの完成形。

MQA が折り畳みとその後デコーダーで用いるロスレスの広げ戻しを達成するためにこの「覆い」の下で行う新たな処理の領域についてはこれらの図には描かれていません。

例えば、サンプリング・カーネルは新たなもので、環境的・音楽的信号傾向に適合させていて、人間の聴感に適応でき、細かな時間領域分解能を持ちながら、微細ビットのロスレス圧縮の解像度を上げています。

MQA ではまた end-to-end 構造の開発でも雑音変調ゼロ化を達成するためデコーダーに減法ディザ技術を導入しており、他方式とは一線を画しています。

MQA の階層性(hierarchy)

MQA はいくつもの方向で階層性があります。

- チャンネル数は 2 以上にスケール可能です。
- 2 倍もしくはそれ以上のレートでサンプルされた音楽では、再生サンプルレートを階層的に最適化可能で、それはプラットフォームに見合った「広げ戻し」で決められます。
- 転送レートは 1 倍、2 倍、4 倍、8 倍にスケール可能です。これまでの我々のやり取りが 1 倍に集中していたのは現在市販されている録音の配信には 1 倍が十分に効率的だったからです。1 倍の転送レートはデータレートが低く、後方互換性が最大です。極端な録音に対しては若干の制約があります。
- アーカイブや高度な配信用には転送レートを 2 倍、4 倍にでき、それに比例して周波数特性と帯域幅は拡張されます。(アーカイブは無制限であるべきですが、

現行再生システムで時間領域精度を保って 50kHz 以上を再生できるものはまだ稀で、研究所レベルです。)

下の表に MQA を総称とする様々なファイル形式をまとめます。MQL ファイルはオリジナルのサンプルレートで、転送レートは最低 2 倍、ロスレス折り畳みは 1 回というもので、カプセル化折り畳みを含むことができます。

Stream	MQA	MQL	MQL	MQL	MQA2x	MQL2x	MQL2x	MQA4x	MQA8x
Original Fs	1x	2x	4x	8x	2x	4x	8x	4x	8x
Transmission Rate	1x	1x	1x	1x	2x	2x	2x	4x	8x
Data Rate Mbps (typ)	1.1	1.2	1.2	1.3	2.1	2.2	2.25	~3.5	~6.5
Audio Folds ¹⁰									
Lossless	√	√	48 kHz	96 & 48 kHz	√	√	96 kHz	√	√
Encapsulation	n/a	n/a	@ 96 kHz	@ 96 kHz	n/a	n/a	@ 192 kHz	n/a	n/a
Precision	24b	24b	24b	24b	24b	24b	24b	24b	24b
Channel Capacity									
Maximum (bits)	23.85	22	22	22	23.85	22	22	23.85	23.85
Typical (bits) _{II}	23.85	>20	18	17.5	23.85	>20	18	23.85	23.85
End-to-end									
@ 20 kHz	0 dB	0 dB	□ .06 dB	□ 0.17 dB	0 dB	0 dB	-0.002 dB	0 dB	0 dB
-1 dB @	n/a	n/a	32 kHz	30 kHz	n/a	n/a	63 kHz	n/a	n/a
-10 dB @	n/a	n/a	52 kHz	54 kHz	n/a	n/a	110 kHz	n/a	n/a
Output format	1x	2x	analogue ¹²	analogue	2x	4x	analogue	4x	8x
Blur (m of air)	10	4	5	4	4	3	3	2	<2
Using	lossless	lossless	Option 8	Option 21	lossless	lossless	Option 5	lossless	lossless

注 11: 典型的音楽録音。表示した数字はチャンネル容量。減法シェーピングによりこの D レンジはオーディオ帯域で 3bit まで良化可能。1 倍 16bit 典型チャンネルの容量は 15.9bit。

注 12: マスタリング技術者は 24kHz 以下のノイズフロアとして通常 19.5~23bit の選択が可能。

注 13: アナログのスペック値は「標準の階層型 DAC」を使用の場合。

注 14: 典型的 192kHz PCM チャンネルでは、等価 post-blur は 10 倍(約 50m の空気)高く、pre-blur は 200~400 μs の範囲。

#

出典: <http://www.stereophile.com/content/mqa-questions-and-answers>