

Introduction

Almost since the arrival of digital audio systems of sufficient quality for the recording, transmission and replay of high-quality sound – and particularly since the introduction of CD in major audio markets during 1982/3 – there have been complaints that digital audio falls short of the perfection it appears to promise. Early problems with the performance of A-to-D and D-to-A converters, excess jitter and inadequate understanding of the need for and correct application of dither all played a part in this, but even as these issues were resolved dissenting voices continued to claim that there was something 'wrong' with digital audio as represented by the 16-bit/44.1 kHz and 16-bit/48 kHz systems prevalent at the time.[1]

It is now widely, although not universally, accepted that 'hi-res' digital audio, with increased sampling rate or bitdepth, delivers improved sound quality. But it does so at large cost to coding efficiency. A 24-bit/88.2 kHz recording requires three times the data rate of a 16-bit/44.1 kHz alternative, and that ratio increases by further factors of two as sampling rate is doubled again to 176.4 kHz and then to 352.8 kHz, the sampling rate of DXD. While the progressive improvement in sound quality is welcome, it takes a disproportionate toll on data rates and storage capacity. Simply increasing sampling rate also fails to address head-on why it is that 44.1 kHz and 48 kHz sampling rates impose subjective limitations. Instead, sampling rate has become a proxy for resolution.

Human Hearing

Academic research has identified two distinct reasons why increasing sampling rate can improve sound quality.

First, despite the well-known c.20 kHz upper frequency limit of the human ear, we are able to perceive ultrasonic frequencies and there is evidence that their presence, at appropriate levels, is pleasurable. [2][3] Many musical instruments generate ultrasonic components at significant amplitude and it appears that this is not circumstantial, as previously assumed, but instead germane to their perception.[4]

Second, recent hearing research provides support for the long-standing notion that the time-domain performance of anti-alias and reconstruction filters – most especially steep digital linear-phase filters – is responsible for perceptible degradation of sound quality. Recently, direct evidence for the audibility of low-pass filters used in digital audio has been published. [5]

It has been known since at least 1946 that the Fourier time-frequency uncertainty inherent in conventional signal analysis can be 'beaten' by human listeners, and by a significant margin. [6][7] Indeed, recent experimental studies have shown temporal discrimination at least 5 times higher. [8][9]

These findings accord with the idea that the capabilities of human hearing have been determined by evolutionary requirements, in particular the need to identify sounds as 'potentially threatening' or 'non-threatening' in the shortest possible time interval, thereby providing the maximum opportunity for fight or flight. While vision plays a part in this too, of course, we cannot see through 360 degrees, around corners, or at as low light levels as some predators.

In these circumstances in particular, our hearing is the primary sense by which we detect danger, and speed of detection and rapid estimation of direction and range is of the essence. As too is the ability to separate direct sound from short-delay or closely-spaced reflections – which naturally require the resolution of short time intervals that are independent of frequency or bandwidth.

Our understanding of natural soundscapes, reverberation, animal vocalisations and speech, requires adjustable time/frequency balances which, up until now, have not been adequately accounted for in audio system design.[10]

This all suggests that the time-domain acuity of the human auditory system has been more important than frequency-domain acuity and explains why its time-frequency uncertainty is so much superior to that of an FFT analyser (and its close relative, the sinc-kernel of digital sampling). Causal signals are key to our achieving this feat; if natural signal waveforms are time-reversed we can no longer outperform the time-frequency uncertainty of Fourier analysis.[11]

Temporal acuity manifests a survival characteristic, one with origins that must reach back to much earlier in the mammalian timeline than the emergence of *homo sapiens*.

It would be strange indeed if our remarkable time-domain acuity were irrelevant to the perception of music. In fact there is persuasive evidence that this is not the case: those experimental subjects who have proven most adept at resolving time-frequency uncertainty are musicians, suggesting that time-domain acuity is enhanced – trained – by the process of becoming a musician.[11] So the traditional frequency-domain view of audio system performance is fundamentally at odds with our perception of music. A fresh approach to the specification and design of high fidelity audio encoding and equipment which takes much closer account of system time-domain performance is therefore long overdue.

A fresh start

MQA (Master Quality Authenticated – the reason for the name will become apparent), embodies just such an approach. As such it isn't just a new technology, it represents a new *philosophy* of high-resolution sound reproduction: one which establishes performance criteria based on the abilities of human hearing and embraces the entire recording/reproduction process from the studio microphone to the replay D-to-A converter in order to meet them. It also has important ramifications for the downstream amplifier and loudspeaker or headphone that complete the chain.

The novelty – indeed, the radicalism – of the MQA approach is summarised in its mission statement: to cause no more 'harm' to an audio signal than its passage through a few metres of air.

In its aim to convey the inherent sound quality of analogue and digital masters with unprecedented fidelity, MQA puts an emphasis on time-domain performance, a process which began with Peter Craven's Meridian-sponsored work on apodising filters.[12]

Concern that the steep low-pass anti-aliasing and reconstruction filters of classical digital audio practice add audible 'time smear' or 'blur' goes back still further, originating in the 1980s, but MQA addresses this concern in a more thoroughgoing and knowledge-based manner than has ever been attempted before. It sets as its target 10μ s time resolution with extant recordings, with a recommendation of 3μ s for future audio archives in order to better the 5-8µs discrimination ability of human hearing. [13]

Conventional 96 kHz recording – for which the sampling period is 10.4μ s – might seem to satisfy the target. However that is to ignore the effect of the anti-alias and reconstruction filters, which employ a sinc kernel to provide inter-sample-temporal resolution – at the expense of elongating the system impulse response far beyond one sampling interval.

To achieve MQA's superior time-domain performance requires knowledge of all the filters present during recording and replay, which is why it is an end-to-end system encompassing the entire chain from the microphone in the recording studio to digital-to-analogue conversion in the home. Even the behaviour of the tape machine is taken into account when legacy analogue material is being encoded.

An MQA decoder is matched to the associated converter while the complementary kernels used in the encoder and decoder combine to give the desired analogue result. Only by controlling the entire process right up to the output of the D-to-A converter in the replay device can MQA's exemplary time domain performance be efficiently guaranteed.[13]

As MQA is also applied to the replay process within the mastering studio, the final sound achieved there is exactly that which is delivered to the home. This can be verified explicitly on replay by, for instance, an LED indictor on MQA-ready hardware, hence the name Master Quality Authenticated: the user is assured that they are receiving exactly the same analogue signal that was signed-off in the mastering process.

MQA's time-domain performance relative to conventional high-quality 24-bit/192 kHz recording-plus-replay is at least an order of magnitude superior. Leading edge uncertainty is reduced from 250 μ s to 3 μ s, total impulse duration from 500 μ s to 25 μ s, and perceptual smear from at least 100 μ s to less than 10 μ s. Overall frequency and impulse responses for MQA are shown in Figure 1, while Figure 2 confirms that MQA's effect is no worse than imposed by sound transmission through a few metres of air.

To put the impulse response of Figure 1 into perspective, Figure 3 compares it to that of a typical 24-bit/192 kHz ADC/DAC chain.

JAS Journal 2015 Vol.55 No.5(9 月号)



Figure 1. End-to-end frequency and impulse responses of MQA transmitted in a 48 kHz channel.



Figure 2. How the impulse response of MQA compares with that of the passage of sound through various distances of air (relative humidity 30 %). (Showing two encapsulation settings).



Figure 3. Comparison of the end-to-end impulse response of MQA compared to that of a typical 192 kHz system.



Figure 4. Effect on frequency and impulse responses of cascading up to eight second-order Butterworth low-pass filters, each with a corner frequency of 30 kHz (such as may occur in a chain of microphone, converters, amplifiers, loudspeaker and so on).

Ancillary Equipment

Such is the improvement in time-domain performance achieved by MQA that analogue devices in the audio chain can be limiting factors in achieving its full subjective benefit. This is the case with most microphones, some electronics and many loudspeakers. The deliberate curtailment of high frequency bandwidth to obviate ultrasonic noise or distortion – previously regarded as best practice by many audio engineers – is deprecated because each low-pass filter downstream of the replay D-to-A converter is uncontrolled and lengthens the system impulse response. Figure 4 illustrates this by showing how impulse response progressively lengthens with up to eight concatenated second-order Butterworth low-pass filters, all having a corner frequency of 30 kHz.[9]

High Resolution Playback

JAS's action plan for High-Resolution Audio, which suggests that amplifiers, loudspeakers and headphones have at least 40 kHz bandwidth, begins the process of ensuring that equipment downstream of the DAC does not limit the sound quality achievable and is to be applauded. [14]

With transducers, wide bandwidth itself isn't sufficient: there must also be no prominent resonances at high frequencies, as is often the case with metal dome tweeters.

JAS Journal 2015 Vol.55 No.5 (9月号)

Convenience versus Quality

Were MQA only an unprecedentedly 'low-blur' recording/replay technology running at a high sampling rate it might be hailed as an audiophile breakthrough but wouldn't have an impact on the sound quality achieved more widely from mobile devices and streaming services.

Figure 5 shows a graph of the notional quality and convenience of different audio carriers, arranged chronologically. Except in the rarefied audiophile arena, sound quality has been progressively overtaken by convenience as the more important factor in widely-experienced music consumption.

MQA removes the compromise between convenience and quality by providing even better sound than is delivered by current 24-bit/96 kHz or 24-bit/192 kHz recordings at significantly reduced data rate. MQA can typically be streamed at ~1 Mbps average – less than the 1.411 Mbps data rate of CD – or replayed over any 24-bit, 44.1 or 48 kHz capable device provided that an MQA decoder is fitted.



Figure 5. Major Label view of how the balance of quality and convenience has changed as new music storage and delivery technologies have been introduced. Also shown is the ideal situation, essentially provided by MQA, in which convenience of distribution, storage and use comes at no cost to sound quality. On the contrary, sound quality is superior to current hi-res digital audio.



Figure 6. Comparison of the data rate of extant coding methods with that of MQA folded to 48 kHz.

An MQA download file, with average data rate of around 1.2Mbps, has full backwards compatibility which means that even if a decoder is absent the file will still play, albeit without the full sound quality enhancement. Even if only a 16-bit 'pipe' is available an MQA decoder will provide superior sound quality to conventional 16-bit 44.1 or 48 kHz replay.

By packaging low-blur audio signals in backwards-compatible 44.1 or 48 kHz files, MQA can go anywhere and be replayed wherever we can play conventional PCM audio files.¹

¹ Note that MQA always maintains the base rate of the digital source signal, so, e.g. 44.1/88.2/176.4/352.8 kHz sources are packaged in a 44.1 kHz container.

Digital Audio Converters and Channels

The closest reproduced sound can come to the original is by connecting a microphone directly to a loudspeaker – this is not perfect, each analogue stage is lossy and indeterminate. To store or transmit that sound, one option is to convert it to digital form, as shown in Figure 7, however very critical steps remain at the analogue-digital and digital-analogue gateways and in the compromises and permanent limitations made at these points.



Figure 7. A conceptual model for recording.



Figure 8. Internal blocks in now widely used delta-sigma A/D and D/A converters, shown connected as PCM.

Figure 8 gives insight to modern converters. The high-speed small-word-size modulators are the critical stages, while for the PCM passed from one to the other, the transmission sample rate may almost be chosen arbitrarily at a fraction of the modulator rate.

As the signal passes through the A/D converter, each stage of down-sampling adds quantization noise and increases temporal blur; while in the D/A converter each corresponding upsampling stage adds further noise and blur. The properties of the decimation and upsampling filters and quality of processing at each stage significantly impacts the overall sound quality. So, the signal chain from the modulator in the A/D to that in the D/A is certainly not lossless – it's not even linear since device-dependent aliasing, thermal and modulation noise and temporal blur are all introduced. The conversion gateways present a major obstacle to transmitting archive-standard analogue audio.



Figure 9. Comparing the temporal blur of conventional sinc-kernel sampling and reconstruction with MQA.

We can see that if we transmit using higher sample rates, then, in typical systems, there are fewer processing stages and less filter smear and these aspects are independent of the benefits of the wider bandwidth potential of higher transmission rates, see Figure 9.

In the perfect world we might directly connect the multi-bit converter modulators to avoid these up- and downsampling stages. By using carefully chosen encoding and decoding filter kernels with matched dither, MQA makes this connection virtually, aiming for the most precise analogue-to-analogue fidelity from the hardware.

Figure 10 compares the coding space for a number of common transmission formats, including CD, and 96 and 192 kHz at 24- and 32-bit depths; also shown is the noise-floor for DSD.

To accommodate the coding capacity of 32-bit LPCM, the vertical scale of this graph is formidable. At the top, 120 dB SPL is the threshold of pain; 0dB (the quietest we can perceive) is in the centre. The noise-floor of the 32-bit channel is 120 dB lower than silence, so although more bits give us a finer staircase, the channel's information capacity is excessive.

Four additional curves provide perspective: red is the noise-spectral human threshold (sounds below this line are not heard directly); green, typical quiet out-door environmental sound; navy, the noise-floor of the quietest commercial recording in our survey and, finally, brown, the fundamental thermal noise limit for a microphone (data below this line are Brownian motion). [15][16]



Figure 10. Shannon diagram showing the capacity of various digital channels (area is equivalent to data rate).



Figure 11. Shannon diagram of an example 24-bit/192kHz recording showing signal and noise levels.

Information Delivery

Figure 11 is a Shannon diagram showing the entire coding space for a 24-bit/192 kHz source signal and, within it, in red, the peak frequency spectrum of real audio – in this case a movement from a Ravel string quartet. The green and blue curves, respectively, show the peak and mean of the recording's background noise – that is the sound we hear before, during and after the music – that is comprised of hall ambience and some analogue noise.

This recording was chosen as a worked example because it is of real instruments in a natural acoustic. The plucked pizzicato strings are challenging to reproduce and the spectrum shows harmonic components well above 20 kHz, which contribute to the sonic envelope.

The peak level collected by the microphone shows a declining spectrum with increasing frequency. This is a very typical feature of naturally-occurring sounds, one which can be exploited to reduce both temporal blur and data rate.[17]

Note that the music and noise curves converge at 'P'-there are probably signal components above that frequency which are lost in the noise.

The region marked 'A' is in the conventional audio band – we are responsive to tones up to 20 kHz. Region 'B' also contains music content, but none of that range is audible if heard in isolation; however elements in 'B' do contribute to the temporal resolution and sonic envelope; experience shows that removing these lowers fidelity. [9]

Region 'C' is different: it carries no salient music-dependent information, is above the passband of almost all microphones and loudspeakers and is also both below and beyond human thresholds for noise signals.

This steady noise is quite inaudible, however the higher sampling rate can enable improved resolution in region 'A', the higher sample rate lowering blur.

As noted earlier, in acoustic recordings, point 'P' is observed between 30 and 60 kHz, with 40 kHz being typical. So, as sample rates are increased above 96 kHz, for example to 192 or 352.8 kHz, region 'C' extends over a wider bandwidth. In that region any music-correlated components are well below the noise, however there can be improved sound reproduction because the higher transmission rate enables less filter blur and better convergence.

So, the orange triangle in Figure 11 encloses all the musically relevant part of the signal (the remainder is noise and silence) and has an area of about one-sixth of the entire coding space – which means that five-sixths of the data rate is squandered.



Figure 12. Application of MLP lossless compression reduces the data rate from 4.6 to 2.9 Mbps.

High-performance lossless compression can improve matters by reducing the data rate to 2.9 Mbps per channel, a saving of 37 per cent, but this is still inefficient and too high to be ideal for streaming from online music services.

The goal of MQA is to deliver the contents of the orange triangle *precisely*, with increased and extreme precision, while avoiding temporal blur or noise modulation in the converters. To achieve this MQA goes 'beyond lossless' in the sense that it has at its core a realisation that 'lossless', as the term is usually used, is no guarantor of ultimate sound quality because it does not embrace the A/D and D/A conversion, volume control etc.

A 16-bit/44.1 kHz digital file can be delivered losslessly but that doesn't ensure that it achieves blameless end-toend sound quality, far from it.

The same is true of conventional hi-res digital because, while potentially superior to 16-bit/44.1 kHz, it has not been specified to match the time-domain acuity of human hearing.

So the goal of MQA is not lossless operation in this narrow, technical sense, even though its core digital path is lossless and bit-for-bit determined and confirmed by the decoder. The goal is to capture and deliver everything we can hear, without the inherent blur of conventional sampling, without the uncertainty of converter quantisations, and to convey the analogue sound heard in the mastering studio to the end-user without modification. This makes it 'lossless' in a much more profound, and relevant, way.[13]

Music Origami

We refer to the process by which MQA reduces high sampling rate signals to $F_s = 44.1$ kHz or 48 kHz as 'music origami', after the Japanese art of paper folding. Figures 14 to 16 show how it works for a 192 kHz source.

Although the process is depicted here in two dimensions, it is actually a three-dimensional construct.

The first 'fold' (Figure 13) reduces the transmission rate from 192 kHz to 96 kHz, and the second (Figure 14) from 96 kHz to 48 kHz. The folding process is not filtering and *the inherent sample rate and bit-depth remain*. In the transport, as we fold, each resulting sample conveys more information. In the MQA lexicon this first folding process is known as 'encapsulation'.

The process is hierarchically scalable, so if the source were, for example, a 352.8 kHz file then we use three folds to reach the final transmission rate of 44.1 kHz. Similarly, if the source were only 96 kHz, then we start with the lossless process of Figure 14. MQA is also hierarchically scalable so that each type of fold can be used one octave higher to enable double-speed transmission options.

The second fold, illustrated by Figures 14 and 15, is able to be lossless because, as illustrated by the noise floor of the original analogue signal (blue trace), much of the available dynamic range with 24-bit encoding is occupied by noise.

So the second fold process uses buried data techniques to reversibly hide the signal information from above 24 kHz within the noise floor below 24 kHz.

Principally because of the combination of environmental noise and microphone self-noise (plus tape noise with analogue masters), very few recordings achieve let alone exceed 16-bit dynamic range. Add to this the fact that we can hear signals within noise only to about 10dB below the noise level (see olive curve in Figure 13) and it follows that bits 19 to 24 carry no useful information.[15]



Figure 13. The first 'fold' is an 'E' or 'encapsulation' fold'. The transmission rate is halved and signal content of C is stored below the noise level in area B.



Figure 14. The second fold is type 'L': the transmission rate is halved again. The signal in area B is buried losslessly and hidden beneath the noise of the baseband spectrum, along with C.

JAS Journal 2015 Vol.55 No.5 (9月号)



Figure 15. A completed MQA download file. If there is a decoder present on replay then the two folding processes are undone precisely. If no decoder is present then the compatible baseband signal is reproduced.

What these diagrams don't convey is the range of novel processes that MQA uses 'under the hood' to achieve this folding and the lossless unfolding subsequently applied by the decoder.

For example, novel sampling kernels are used which suit environmental and music signal statistics, which are adapted for human listeners and provide tight time resolution, while fractional-bit lossless coding increases resolution.[19][10]

MQA also exploits its end-to-end architecture to enable subtractive dither techniques in the decoder to ensure that, unlike other formats, there is zero noise modulation. Unlike lossy codecs, MQA never ever intrudes upon the music signal, plus it maintains precise and constant characteristics throughout an entire musical work.[20]

Playback

If an MQA decoder is present then these two folding processes are undone to restore, in this case, a 24-bit/192 kHz output with the end-to-end frequency and impulse responses already shown in Figure 1. Furthermore, the decoder reconstructs the signals fed to the D/A converter with bit-accuracy, giving the identical result heard in the mastering studio; it authenticates and indicates this result.

If no decoder is present then the file replays at 44.1 kHz or 48 kHz, providing backwards compatibility.

If the transmission path can only pass 16 (rather than 24) bits, then an MQA decoder can reconstruct the baseband (area A in Figure 10) along with a lossy version of B (and C), which sounds very close to the high-res original and much better than CD replay.

Similarly, if the replay device is not capable of the highest sampling rate, the unfolding process can be stopped part-way through and the decoder optimises the temporal and frequency response.

Summary

MQA addresses two traditionally incompatible needs: quality and convenience. It provides sound quality which significantly exceeds that of current hi-res digital formats but it does not require high data rates, nor does it disenfranchise users who don't own a decoder. Only one release format is required to cover the gamut of replay possibilities, and MQA material can be distributed like conventional PCM: streamed, made available as a file download, even put on a CD or other optical disc.

MQA has been demonstrated to many of the world's leading recording engineers, music producers and record companies and has been enthusiastically received. Demonstrations to the press and public have resulted in widespread acclaim for it representing the future of digital audio.

Further Reading

A wide-ranging overview of the background to MQA is available in an Audio Engineering Society paper delivered at the 137th Convention in October 2014.[13]

Patent Notice

Several aspects of MQA are subject to patents granted and pending.

References

- [1] ADA, 'Proposal of Desirable Requirements for the Next Generation's Digital Audio', Advanced Digital Audio Conference, Japan Audio Society (April 1996)
- [2] Oohashi, T., et al, 'Inaudible High-Frequency Sounds Affect Brain Activity: Hypersonic Effect', J. Neurophysiol, 83: 3548–3558, (2000).
- [3] Oohashi, T., et al, 'Multidisciplinary study on the hypersonic effect', International Congress Series 1226, pp 27–42, (2002). <u>http://dx.doi.org/10.1016/S0531-5131(01)00494-0</u>
- [4] Boyk, J., 'There's life above 20 kilohertz! A survey of musical instrument spectra to 102.4 kHz', http://www.cco.caltech.edu/~boyk/spectra/spectra.htm (2000)
- [5] Jackson, H. M., Capp, M. D. and Stuart, J. R., 'The audibility of typical digital audio filters in a high-fidelity playback system', Convention Paper 9174, 137th Audio Engineering Society Convention, (2014).
- [6] Gabor, D., 'Theory of Communication', *Journal of the Institution of Electrical Engineers*, 93, III, p. 429, (November 1946). <u>http://dx.doi.org/10.1049/ji-1.1947.0015</u>
- [7] Gabor, D., 'Acoustical Quanta and the Theory of Hearing', *Nature*, **159**, pp. 591–594, (1947). <u>http://dx.doi.org/10.1038/159591a0</u>
- [8] Oppenheim, J. M. and Magnasco, M. O., 'Human Time-Frequency Acuity Beats the Fourier Uncertainty Principle', *Phys. Rev. Lett.*, **110**, 044301, (2013). <u>http://dx.doi.org/10.1103/PhysRevLett.110.044301</u>
- [9] Oppenheim, J. M., et al., 'Minimal Bounds on Nonlinearity in Auditory Processing' (Jan 2013). arXiv:1301.0513 q-bio.NC.
- [10] Lewicki, M.S. 'Efficient Coding of natural sounds', *Nature Neurosci.* 5, 356–363 (2002). http://dx.doi.org/10.1038/nn831
- [11] Oppenheim, J. M., et al, 'Degraded Time-Frequency Acuity to Time-Reversed Notes', PLOS ONE, 8, pp. 1–6 (June 2013). <u>http://dx.doi.org/10.1371/journal.pone.0065386</u>
- [12] Craven, P.G., 'Antialias Filters and System Transient Response at High Sample Rates', J. Audio Eng. Soc., Vol. 52, No. 3, pp. 216–242, (March 2004)
- [13] Stuart, J. R. and Craven, P., 'A Hierarchical Approach to Archiving and Distribution', Paper 9178, 137th Audio Engineering Society Convention, (2014).
- [14] Japan Audio Society (JAS), 'JAS's action plan for High-Resolution Audio', (June 2014)
- [15] Stuart, J.R., 'Noise: Methods for Estimating Detectability and Threshold', J. Audio Eng. Soc., 42, 124–140 (March 1994)
- [16] Fellgett, P.B., 'Thermal noise limits of Microphones', J. IERE, **57** No. 4, 161–166 (1987). http://dx.doi.org/10.1049/jiere.1987.0058
- [17] Dragotti, P.L., Vetterli, M., Blu, T., 'Sampling Moments and Reconstructing Signals With Finite Rate of Innovation: Shannon Meets Strang-Fix', *IEEE Trans.Sig. Proc.* Vol. 55, No. 5, pp. 1741–1757, (May 2007) <u>10.1109/TSP.2006.890907</u>
- [18] Rieke, F., et al, Spikes: Exploring the Neural Code, ISBN 978-0-262-18174-7, MIT Press (1997)
- [19] Unser, M. 'Sampling 50 Years after Shannon', Proc. IEEE vol. 88 No. 4, pp. 569–587 (Apr. 2000). <u>http://dx.doi.org/10.1109/5.843002</u>
- [20] Widrow, B., Kollár, I., Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications, CUP, Cambridge, UK, ISBN: 0521886716 (2008)