UNIVERSITY OF
FORWARD
THINKING
WESTMINSTER⌗

WestminsterResearch
http://www.westminster.ac.uk/westminsterresearch

Evaluating analog reconstruction performance of transient digital audio workstation signals at high- and standard- resolution sample frequencies

Toulson, R.

This is an accepted manuscript of a book chapter published by Routledge in Innovation in Music: Performance, Production, Technology, and Business on 23 April 2018, available online:

http://www.routledge.com/9781138498198

Evaluating analog reconstruction performance of transient digital audio workstation signals at high- and standard- resolution sample frequencies

Rob Toulson

University of Westminster, London, UK

r.toulson@westminster.ac.uk

Abstract

Given that audio signals in many applications are neither predictable nor guaranteed to be repeated (and hence do not deliver an infinite history of data), they do not completely adhere to the ideal sampling theorems presented by Nyquist and Shannon. Digital-to-analogue (DAC) reconstruction theory is hence used to investigate how accurately digital audio workstation signals are actually reconstructed, and explores whether high-resolution sample frequencies (i.e. frequencies above 44.1 kHz) provide a performance advantage. Ideal reconstruction profiles are then evaluated against the actual reconstruction data observed from three pro-audio DACs and at multiple sample frequencies. The test signal is chosen to evaluate the performance of DACs when presented with transient data that approaches the Nyquist sample frequency. This approach is used because it has the potential to yield information on the suggested benefits of higher-than-Nyquist sample and reconstruction approaches in a real-world music production context.

1. Introduction

High-resolution (or hi-res) audio can be defined as digital audio data that has greater amplitude resolution than 16-bit or greater time-axis resolution than 44.1 kHz (Rumsey 2007). As the compact disc delivery format itself delivers 16-bit and 44.1 kHz accuracy, hi-res can sometimes be described simply as "greater than CD" resolution. In modern music production projects, 24-bit recording and reproduction is standard, but considerations between 44.1 kHz, 96 kHz and 192 kHz sampling rates are still made by many professionals.

The performance of digital audio sampling continues to generate debate amongst experts, even given the thoroughly documented theories developed first by Nyquist (Nyquist 1924) and Shannon (1949). Rumsey (2004: 34-35) reflected on this debate, reporting that strong arguments remain for sampling at no higher than a rate at double the bandwidth of human hearing, as specified by Nyquist's and Shannon's sampling theories. Conversely, Rumsey also notes that many professional studio engineers and music producers are in favour of higher sample rates in practice.

Psychoacoustics research by Moylan (1987) indicated that listeners can hear the difference between audio sampled at higher resolutions than the Nyquist frequency for the threshold of human hearing (40 kHz), particularly with the onset of transients. Higher sampled audio is described as sounding warmer, sweeter and fuller. More specifically, Moylan experimented and deduced that humans can hear the influence of a 45 kHz frequency superimposed on a 15 kHz fundamental, even though humans cannot

hear the 45 kHz frequency when it is emitted alone. Moylan concluded that humans can hear above the standard 20 kHz range in more complex sounds, though not specifically on single sinusoids. A significant challenge to Moylan's results is that the rationale for subjects hearing the additional ultrasonic components is owing to intermodulation distortion caused by non-perfect playback electronics and hence additional distortion components present within the audible (i.e. sub-20 kHz) range, as discussed by Colletti (2013) and Lavry (2012).

A number of listening tests have been conducted to help decide if high-resolution audio is identified and perceived as an improvement by listeners. Jackson *et al* (2014) evaluated the audibility of digital audio filters in order to identify if every aspect of audio signals can be conveyed using only frequencies below the Nyquist limit of a standard CD recording (i.e. 22.05 kHz). Their conclusion was that audible signals do exist that cannot be encoded transparently by a standard CD. Reiss (2016) gives an overview and meta-analysis of published psychoacoustic testing for evaluating audio signals sampled at higher than the compact disc standard of 44.1 kHz. Reiss' study observed the findings of 12,000 different individual listening tests and draws the conclusion that high-resolution audio has a small but important advantage in its quality of reproduction over standard-resolution audio content.

Stuart (2015) makes a number of important observations and claims that are under-investigated, arguing that 'the sounds that are important to us' are not represented by the frameworks described by Nyquist and Shannon. Stuart emphasises that sound is not inherently band-limited and hence there are existent audio signals that extend beyond

the documented human threshold of hearing. Equally, Stuart explains that music and audio sounds do not have an infinite nature, a predictable occurrence or repetition, which are all factors that the conventional signal reconstruction theories rely on.

In this paper, we look specifically at analyses of Shannon's reconstruction theories for common audio signals that do not adhere to the ideal framework – signals which are neither infinitely repeating nor predictable in occurrence. The evaluation is intended to explore how the ideal reconstruction theory performs when considering data that more closely represents produced music than bandlimited infinite sine waves.

2. Audio Signal Reconstruction and Modern Music Production

The Whittaker-Shannon Interpolation Formula (WSIF), - sometimes called the Ideal Interpolation Formula or the Sinc Interpolation Formula - states that a band-limited continuous-time signal x(t) of bandwidth B Hertz can be discretely sampled and uniquely recovered by Equation 1, providing that the sampling rate Fs>2B (Shannon 1949). The WSIF is expressed as

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT) \frac{sin\left[\left(\frac{\pi}{T}\right)(t-nT)\right]}{\left(\frac{\pi}{T}\right)(t-nT)}$$

(1)

where the continuous time signal x(t) is sampled at discrete nT intervals. T is the sampling period given by T=1/Fs.

The WSIF assumes an infinite number of samples and an infinitely repeating time signal, as well as a history of infinite data (as shown in Equation 1, data for all values of n between -∞ and +∞). Hence, the WSIF is "non-causal and physically non-realisable" (Proakis and Manolakis 1992: 425).

Audio signals in many applications are neither predictable nor guaranteed to be repeated, and hence do not deliver an infinite history of data. Real audio signals therefore do not guarantee to obey the concepts of ideal signal reconstruction at all sample frequencies. Despite this, reconstruction techniques are rarely evaluated for transient signals at sample frequencies nearing the Nyquist limit. Given the continuing debates regarding hi-res audio, it is therefore important to model and evaluate the reality of signal reconstruction with respect to transient and non-sinusoidal signals, in order to acquire a better representation of how signal reconstruction performs in practical audio systems. If it has been agreed that the whole bandwidth of $20 - 20,000$ Hz should be uniformly reconstructable for high-fidelity audio, then the reconstruction of a signal containing 20 kHz content should be equally as successful as that of a 100 Hz signal, irrespective of whether listeners can generally perceive a difference or not. The arguments for standard-res audio sampling being sufficient are based substantially on Shannon's theories and the ability for the WSIF to perfectly reproduce sampled analogue signals.

In a practical context, and with reference to music production, modern digital audio workstations (DAWs) and associated processing tools (plugins) have the ability to create and manipulate signals with no regard for standard sampling and reconstruction

criterion. For example, a transient percussion signal might be sampled at 44.1 kHz through a 22 kHz band-limiting (anti-aliasing) filter, yet, once in the digital domain, it might be processed with digital dynamic range or wave-shaping tools that attempt to extend the theoretical bandwidth beyond the previously enforced Nyquist limit. A second example is a software synthesizer tool that attempts to create a square wave output, which would theoretically incorporate unlimited odd harmonics. The audio signal presented to the reconstruction filter within the digital-to-analogue convertor (DAC) is digitally manipulated post-antialiasing, so the performance of reconstruction could become unpredictable. In this scenario, it may be that the reconstruction filter is simply not capable of realising the effect of non-linear processing tools that have been introduced, or the reconstruction of such processed signals might introduce artefacts that could become audible. Additionally, in music production (particularly scenarios utilising multitrack synchronised audio), temporal accuracy is of significant importance. Indeed, temporal errors in reconstruction may be more audible and subjectively detrimental than artefacts and distortions identified and measured in the frequency domain.

Conventional testing and analysis of audio systems and signal processes is usually conducted with a continuous 1 kHz test frequency. However, in order to evaluate the reconstruction of DAW processed audio signals and potential benefits of utilising high-resolution sample rates, the WSIF's performance for all audio signal types, should be considered in more detail than has previously been conducted in prior published

research and analyses. Of particularly interest for modelling and analysis are transient waveforms and those with fundamental frequencies approaching the Nyquist limit.

3. Evaluating the Reconstruction of Transient Audio Signals

In the following modelling experiment, an 8 kHz sine wave with a transient decay is chosen for evaluation, as shown in Figure 1.
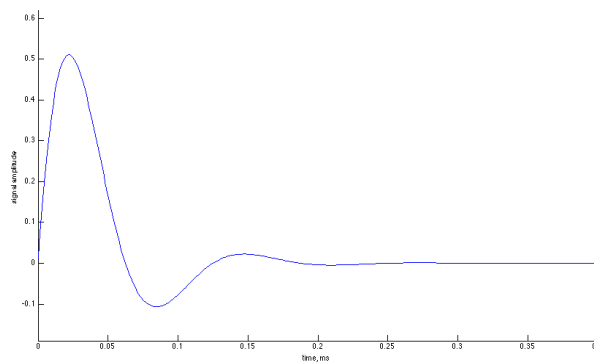

Figure 1. Test signal for transient reconstruction analysis.

The 8 kHz sine wave is chosen because it is a suitably high frequency signal that comes sufficiently close to the standard (44.1 kHz) Nyquist sampling limit, whilst still being audible to healthy listeners. The 8 kHz frequency is also a significant frequency in music production, and manipulation of this frequency with equalisation can make a substantial difference to the audible attributes of produced music; indeed many music producers recommend manipulating this frequency to reduce sibilance in a singer's voice or to add clarity, presence and sparkle to recorded instruments (Izhaki 2008: 212-213; Owsinski 2013: 27). The test signal was created with a 1 MHz sample frequency in

order to - as close as possible - mimic an analogue signal that can be sampled and reproduced at a number of reconstruction frequencies.
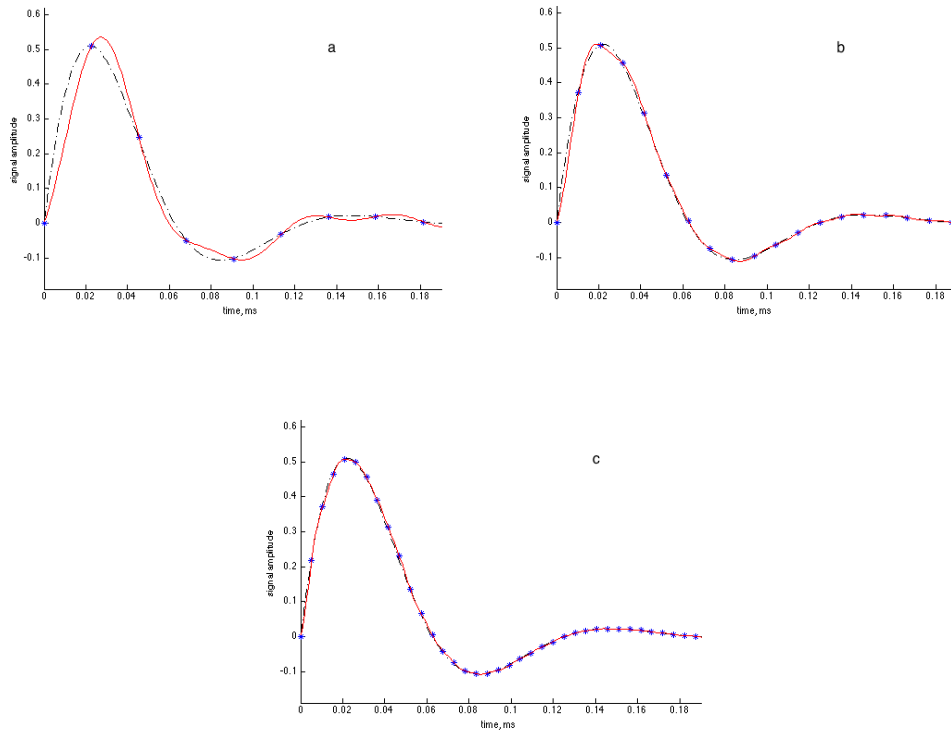

Figure 2. Reconstruction profiles for an 8 kHz transient sinusoid sampled at (a) 44.1 kHz, (b) 96 kHz (c) 192 kHz. Reconstruction signal shown as solid line with analogue waveform as dash-dot line. Sample points shown as asterisks.

Figure 2 shows the success of the WSIF reconstruction for the finite decaying 8 kHz sinusoid signal sampled at Fs values of 44.1 kHz, 96 kHz and 192 kHz. It can be seen that at 192 kHz, the reconstructed signal is superimposed almost exactly over the original continuous-time signal and the sampled data. With 44.1 kHz sampling and

reconstruction, discrepancies between the original signal and the reconstructed signal can be seen.

Most notably, the 44.1 kHz reconstruction is unable to reproduce the rapid attack profile of the transient signal. Its temporal profile has been altered, in that the attack onset is at a reduced gradient and the peak value occurs later than that of the original signal. This temporal error could potentially impact on the accuracy and synchronicity of audio playback, particularly in a multichannel setup, and alter subjective musical attributes such as perceived "tightness" and "crispness".

Figure 3 shows the Fast Fourier Transform (FFT) frequency spectra of the 1 MHz sampled test signal, as well as the frequency profiles of the reconstruction data shown in Figure 2. Each spectra in Figure 3 shows the broad fundamental 8 kHz peak of the test signal. Given the transient nature of the signal and the small number of oscillations before the test signal decays to zero, a broadband peak is expected from the FFT calculation, and it can be seen that the 8 kHz peak occupies all of the audible range up to and above 20 kHz.
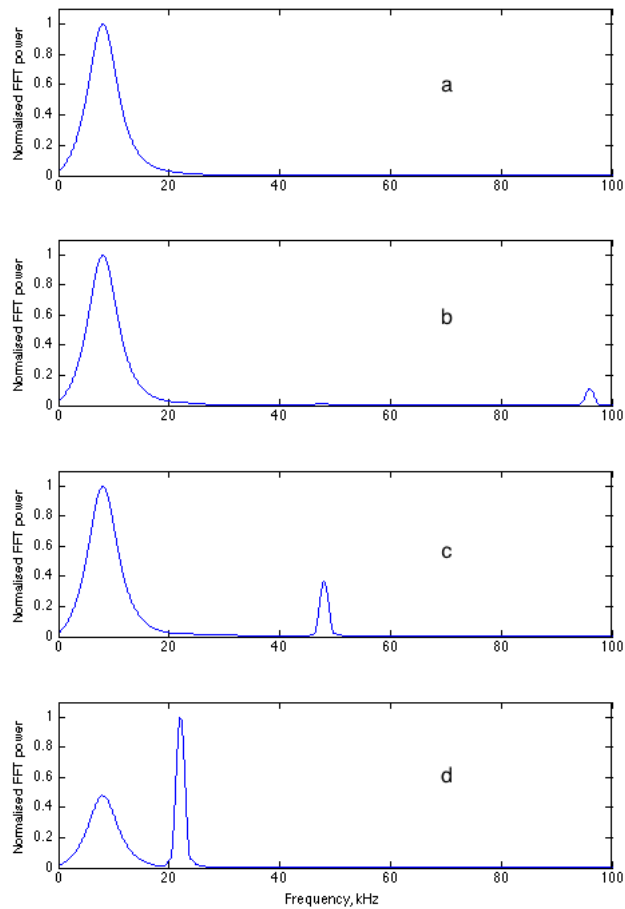
Figure 3. Frequency spectra of the 1 MHz sampled test signal (a) and the reconstruction models calculated for (b) 192 kHz, (c) 96 kHz (d) 44.1 kHz sample frequencies.

It can be seen that the WSIF reconstruction for discrete samples introduces an alias peak at the Nyquist frequency for each of the three sample frequency models. These aliases are at high (inaudible) frequencies for 192 kHz and 96 kHz models, but the 44.1 kHz alias is seen to be significantly powerful and close to the human threshold of hearing. While these artefacts are anticipated to be inaudible to human listeners, there is a

possibility that intermodulation distortion components, within the audible range, could be introduced in an onward processing system as a result of errors in the transient reconstruction.

To evaluate the WSIF as the sampling ratio approaches that of the Nyquist limit, i.e. at Fs/B=2.2, a 20 kHz transient test signal is also considered with a sample frequency of 44.1 kHz (shown in Figure 4). It can be seen that the reconstructed signal is substantially different from that of the original time-domain profile, meaning that accurate reconstruction is not achieved for this transient signal with a fundamental frequency close to the Nyquist limit.
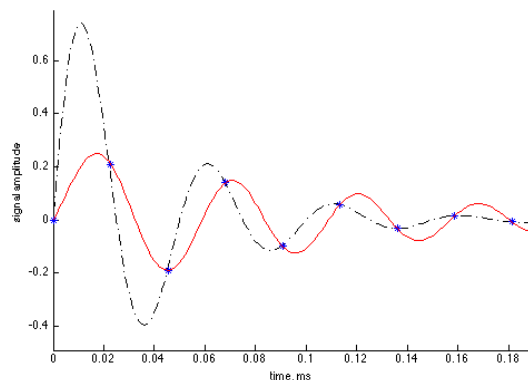

Figure 4. Reconstruction profile for a 20 kHz transient sine wave sampled at 44.1 kHz. Reconstruction shown as solid line with analogue waveform as dash-dot line. Sample points shown as asterisk.

It is therefore shown that the WSIF, i.e. the transfer function of an ideal reconstruction filter, does not uniformly reconstruct signals for the entire Nyquist bandwidth if the signal is transient and made up of a finite number of data samples. The analysis

indicates that, when considering transient signals, higher sample rates above the Nyquist criterion (i.e. at $> 2*B$) are required for accurate reconstruction of the entire audible (20 – 20,000 Hz) frequency range.

4. Hardware Verification

To verify the accuracy of the WSIF and transient signal reconstruction in practical digital-to-analogue audio conversion systems, a hardware verification exercise was conducted. Digital signals were generated at the designated Fs rates in Matlab and rendered as Microsoft Wave (.wav) audio files. These test wave files were loaded into Fidelia audio playback software and output from a number of different audio interfaces, including TC Electronic Studio Konnekt 48, Focusrite Saffire 6 and Cambridge Audio DacMagic Plus devices. Analogue signal profiles were captured with a GW Instek digital oscilloscope utilising a 5 MHz sample rate.

For each hardware interface, the transient test signal shown in Figure 1 was captured after analogue reconstruction for 44.1 kHz, 96 kHz and 192 kHz sample frequencies. Time and frequency domain results for signal reconstruction with the TC Electronic Studio Konnekt, Cambridge DacMagic and Focusrite Saffire are shown in Figures 5, 6 and 7 respectively.

Looking first at the temporal performance of the TC Electronic DAC (Figure 5), it is seen that the 192 kHz reconstruction most closely resembles the profile of the test signal shown in Figure 1. The 96 kHz reconstruction is very similar, whereas the 44.1 kHz reconstruction shows a much reduced attack gradient and hence a greater period

between the onset of the transient and the peak value, which corroborates properties of the 44.1 kHz model result shown in Figure 2a. The frequency domain data in Figure 5 shows no significant artefacts for the 192 kHz and 96 kHz reconstructions, though the 44.1 kHz reconstruction does show evidence of the 22 kHz artefact that is seen in the model results shown in Figure 3d.
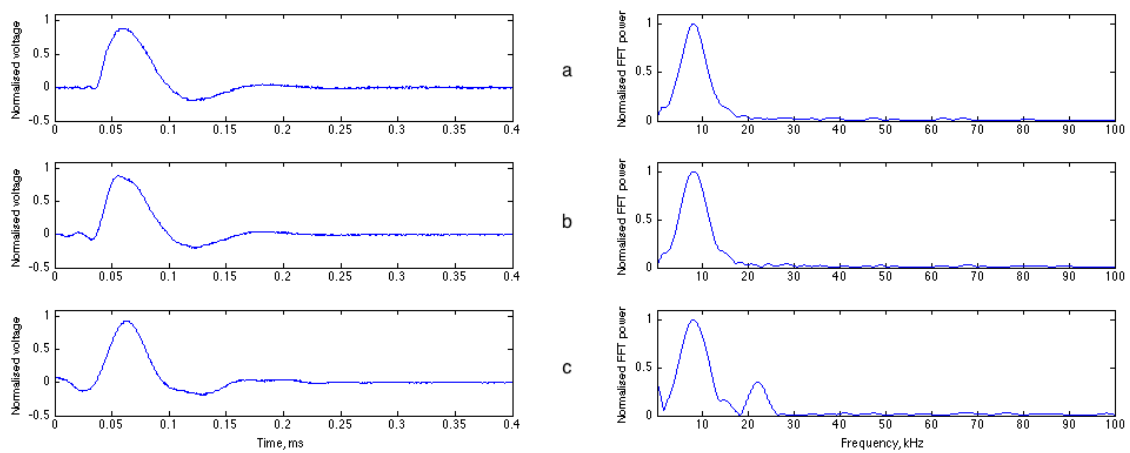

Figure 5. TC Electronic Studio Konnekt reconstruction of transient waveform

data at Fs = 192 kHz (a), 96 kHz (b) and 44.1 kHz (c).

The Cambridge DacMagic reconstruction profiles (Figure 6) similarly show a most accurate time domain reconstruction for the 192 kHz sample rate. Again, the 44.1 kHz reconstruction has a significantly reduced attack gradient and hence a greater period between the onset of the transient and the peak value. The 96 kHz and 44.1 kHz signals show clear sinusoid artefacts on the waveform, which are evident in the frequency domain plots and matching those shown in the model results of Figures 3c and 3d. The

artefact frequency peak of the 44.1 kHz reconstruction is seen to bleed significantly into the audible (sub-20 kHz) range.
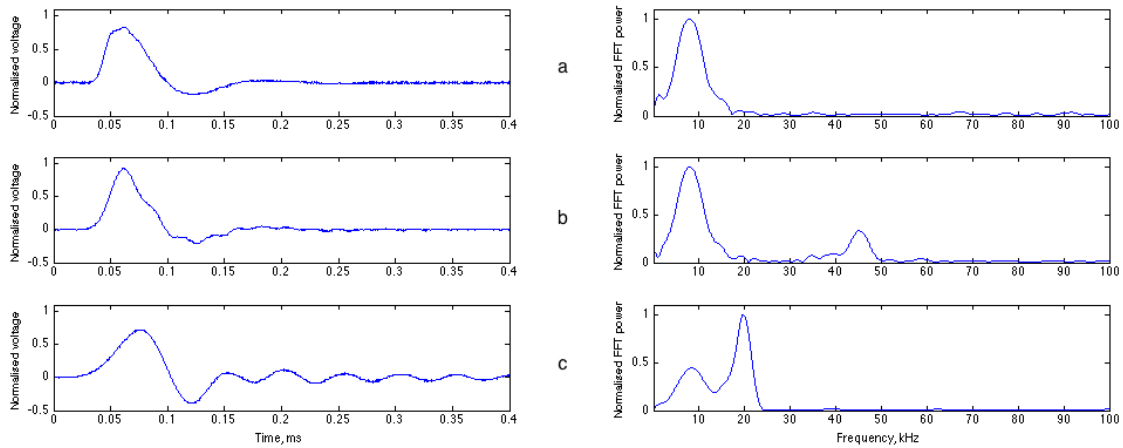

Figure 6. Cambridge DacMagic Plus reconstruction of transient waveform data at Fs = 192 kHz (a), 96 kHz (b) and 44.1 kHz (c).

The Focusrite Saffire (Figure 7) also shows significantly better reconstruction performance at 192 kHz, though the 96 kHz reconstruction also closely resembles the ideal signal profile shown in Figure 1. The Focusrite DAC is USB powered and unsurprisingly displays a lower signal-to-noise ratio, which can be seen on both the time domain signal and as spurious data on the frequency domain plot. As with all the DACs tested, the 44.1 kHz reconstruction has a significantly reduced attack gradient and hence a greater period between the onset of the transient and the peak value. The frequency domain data for the 44.1 kHz reconstruction again shows the alias component as seen in the model shown in Figure 3d. Here it is also seen that the 8 kHz test frequency is significantly altered and corrupted in the reconstruction.
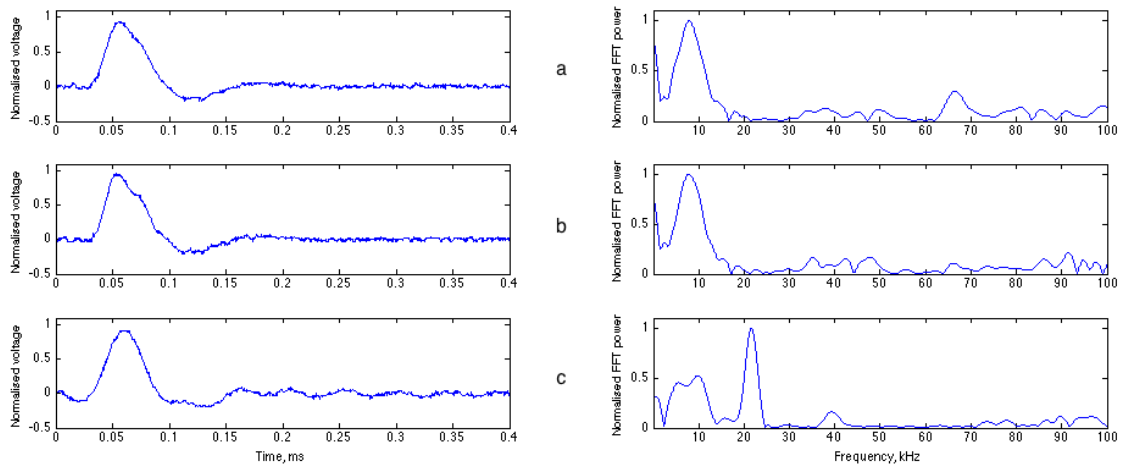
Figure 7. Focusrite Saffire 6 reconstruction of transient waveform data at Fs = 192 kHz (a), 96 kHz (b) and 44.1 kHz (c).

The purpose of this analysis is not specifically to evaluate system performance between devices, but moreover to verify the WSIF model results and identify if higher sample frequencies yield performance benefits in terms of reconstruction accuracy. In all cases the 192 kHz reconstruction shows greater temporal accuracy and less introduced artefact in the frequency domain.

5. Discussion, Conclusions and Future Work

It is shown by the presented analyses that, in practice, the WSIF becomes compromised for high frequency signals which are transient and non-infinite in repetition – i.e. signals that could be described as authentic produced audio or music signals, or components thereof. In particular, transient signals, such as those associated with percussion instruments, can therefore expect to be more authentically reproduced when stored and replayed digitally at higher sample frequencies. It is seen in both the mathematical

models and the hardware testing that the transient signals reproduced at lower sample frequencies have more shallow gradients of attack with delayed positioning of the transient peak, as well as having potentially audible artefacts introduced in the frequency domain. This is generally owing to the fact that in order to perform accurately, the WSIF requires a pre-filled history of data that perfectly reflects the signal to be reconstruction, which is not likely to be the case for real audio and music signals that are reconstructed through a non-ideal (i.e. non-brickwall) analogue filter. In the hardware DAC test results, differences were seen between the types of artefacts and errors of reconstruction at 44.1 kHz, most likely owing to subtly different reconstruction filter designs being implemented by different manufacturers. All hardware DACs, however, reproduced the test signal very well at 192 kHz.

In practice, transient audio signals are generally unpredictable and encountered after a brief period of silence, especially those relating to percussion signals such as kick and snare drums, so temporal accuracy is of importance when considering music production projects and particularly multitrack audio that relies on precise synchronisation of audio data. It is shown in this investigation that the time-domain accuracy of reconstruction is reduced for transient signals that have frequency components approaching the Nyquist limit, so raising the Nyquist limit (by implementing a higher sample frequency) can be beneficial in practical scenarios.

As Stuart (2015) emphasizes, audio signals are not inherently band-limited and, even if a band-limiting (anti-aliasing) filter is used at the point of digital conversion, once in the digital domain today's music producer has many tools to manipulate the waveform

samples regardless of ideal sampling theories. For example, compression and distortion tools can easily turn pure sine waves into hard-clipped waveforms; transient envelope tools can sharpen the attack profiles of percussion transients; and synthesizer tools can generate harsh square wave signals with full-scale slew between individual samples. If the authenticity of the reconstruction is desired to be most accurate, then the use of higher sample frequencies, as shown by this research, can be expected to give a performance advantage.

The analysis presented here therefore, in many ways, aligns with the perception of critical listeners in the music production industries who claim that transients and temporal accuracy are audibly improved with high-resolution audio utilising sample frequencies at 96 kHz (4.8 times the audible frequency bandwidth) and 192 kHz (9.6 times the audible bandwidth).

It is common to quantify the performance of audio systems by electronic measurements, i.e. rather than through subjective listening. For example, total harmonic distortion is regularly quantified by electronic signal analysis with a test signal at 1 kHz. Such quantified measurements are used to give an indication of the performance or "quality" of the audio device and an assumption is made that the performance parameters correlate with the subjective listening experience. With this regard, it is therefore suggested that quantified performance metrics could be developed for measuring the capability of DACs to accurately reproduce digital waveforms at standard and high-resolution sample frequencies – particularly transient test signals as opposed to continuous sinusoidal signals. While it is necessary to evaluate performance at the

standard 1 kHz test measurement, it is suggested that all audio equipment should also be graded on performance for signals close to the Nyquist limit, and particularly with respect to the reproduction and signal processing of transient signals.

The results gathered show that the WSIF model and the actual DAC reconstruction of an 8 kHz transient signal is more accurate when using a 192 kHz sample frequency as opposed to a 44.1 kHz sample frequency. It is not proven, however, whether healthy listeners are able to repeatedly identify this performance improvement in listening tests, though analysis has shown that temporal errors and spectral artefacts within the human hearing range are potentially generated. It could therefore be hypothesized that, for some test material, a difference could feasibly be identified by listeners. This hypothesis will form the basis for future testing.

In general, there are still two fundamental unanswered questions with regards to the human perception of high-resolution audio, those being:

1. Can ultrasonic (i.e. >20 kHz) audio components and qualities be perceived when using high-resolution audio sample rates?

2. Can high-resolution sample rates noticeably improve the reproduction quality of 20 kHz band-limited audio?

These two questions will be considered in future experimentation related to the research results presented here and with respect to common processes in music production. In particular, the design of listening test experiments are critical in achieving conclusive results. When considering listening tests with produced music material, it is important

to ensure that the entire music production signal chain extends to suitably high ratings, for example using ultrasonic microphones, suitably high sample rates, anti-aliasing filters and low-distortion hardware to avoid perceivable sub-20 kHz intermodulation artefacts. These testing methods will be evaluated and perfected in future studies in order to confirm whether the theoretical and measured advantages of high-resolution audio shown in this paper can be perceived by listeners in both laboratory and social listening scenarios.

6. References

Colletti, J. (2013) "The Science of Sample Rates (When Higher Is Better — And When It Isn't)," *Trust Me I'm a Scientist (website),* available online from http://www.trustmeimascientist.com/2013/02/04/the-science-of-sample-rates-when-higher-is-better-and-when-it-isnt/

Izhaki, R. (2008) *Mixing Audio*, 2nd edition, Oxford, Taylor and Francis, pp. 212-213.

Jackson, H. M., Capp, M. D. and Stuart, J. R. (2014) "The audibility of typical digital audio filters in a high-fidelity playback system," *Proceedings of the 137th Audio Engineering Society Convention*, Los Angeles.

Lavry, D. (2012) "The Optimal Sample Rate for Quality Audio," *Lavry Engineering Inc. (website)*, available online at http://www.lavryengineering.com/pdfs/lavry-white-paper-the_optimal_sample_rate_for_quality_audio.pdf

Moylan, W. (1987) "A systematic method for the aural analysis of sound in audio reproduction/reinforcement, communications and musical contexts," *Proceedings of the 83rd Convention of the Audio Engineering Society*, New York.

Nyquist, H. (2007) "Certain factors affecting telegraph speed," *Bell Systems Technical Journal*, Vol. 3, p. 324.

Owsinski, B. (213) *The Mixing Engineer's Handbook*, 3rd edition, Boston, Cengage Learning, p. 27.

Proakis, J. G. and Manolakis, D. G. (1992) *Digital Signal Processing: Principles, Algorithms and Applications*, 2nd edition, New York, Macmillan Publishing Company, p. 425.

Reiss, J. D. (2016) "A meta-analysis of high resolution audio perceptual evaluation," *Journal of the Audio Engineering Society*, Vol. 64, No. 6, pp. 364-379.

Rumsey, F. (2004). *Desktop Audio Technology: Digital Audio and MIDI Principles*, Oxford, Focal Press, pp. 34-36.

Rumsey, F. (2007) "High Resolution Audio," *Journal of the Audio Engineering Society*, Vol 55, No 12, pp. 1161-1167.

Shannon, C. E. (1949) "Communication in the presence of noise," *Proceedings of the Institute of Radio Engineers*, Vol 37, p. 10–21.

Stuart, B. (2015) "High Resolution Audio: A perspective," *Journal of the Audio Engineering Society*, Vol. 63, No. 10, pp. 831-832.