Letter to the editor

Odiviu Popa, Richmond Hill, Ontario,Canada, writes:

Dear Editor,

I have read with great interest Mr. van Maanen's article "On the audibility of "high resolution" digital audio formats and how to test this" in Linear Audio Vol. 5. Such topics are of great value in (hopefully) clearing the muddy waters that are surrounding the subjective evaluation of audio quality. As such, any new drop in the bucket helps better understanding on how to reach the ultimate Nirvana of audio reproduction.

Before going into the details, I did some sanity checks on the fundamentals of this article and, I have to admit, the prospects were not very good. Let's start with section 4. "The historical background…" The author makes a bold statement "One of the underlying requirements for the application of the Fourier theory is that it is applied to linear systems". Whatever way I am looking at this statement, it is incorrect. A function (or system, if you like) f(x) subject to Fourier transform has to comply only to the Dirichlet conditions: 1. the function must be periodic; 2. it must be single-valued and continuous, except possibly at a finite number of finite discontinuities; 3. it must have only a finite number of maxima and minima within one periodic; 4. the integral over one period of |f(x)| must converge. If the above conditions are satisfied, then the Fourier series converge to f(x) at all points where f(x) is continuous. As you see, there is absolutely no requirement or mentioning about the f(x) linearity. Now, the Fourier transform itself is a linear operator, but I believe this has no bearing on what the author was referring to here.

Hans van Maanen replies:

Thank you Ovidiu for commenting on my article. I will reply to each of your points in turn, which will hopefully help to place the various arguments in perspective. In my understanding there is a strict, distinct and clear difference between a "system" and a "function". It is therefore obvious that substituting "function" where I write (and thus mean!) "System" will lead to a misinterpretation of the I message I try to convey. With a "system", let's call it P(ω), I mean an entity which reacts to an input "function" (a real valued function of time, fulfilling the criteria to be Fourier Transformable), let's call it f(t)) to produce an output response (which is also a function of time), which we can call h(t). When the system P(ω) is linear, this means that if the input function is A·f(t) with A being a real-valued scalar of any value, the output function is A·h(t) for all functions f(t) fulfilling the criteria mentioned above. An example is a passive cross-over filter: when one wants to predict the response function h(t) to an input function f(t), this can be done by the application of Fourier theory. However, when the system is non-linear, this will lead to erroneous results if the non-linear properties of the system are ignored. Surprisingly, everybody accepts that e.g. an amplifier will produce a more or less distorted output signal, the distortion depending on the properties of the input signal (like its temporal properties, signal strength, etc.) but most people think that the output of our hearing is, except for the strength, independent of the properties of the incoming sound as if it were a linear system. So by recording the incoming sound with a linear transducer (e.g. a high quality microphone) they think that they can *predict what we will hear. But as our hearing is a strongly non-linear transducer, this is an incorrect application of the Fourier theory and will thus lead to incorrect conclusions.*

Going on now to Fig. 4 and Fig. 5, the author states that the signals have "exactly the same spectral composition". It is easy to prove this is wrong. First, the signal in Fig. 4 is not directly Fourier transformable; as it is, it violates Dirichlet condition 1 (periodicity). Now, it is not difficult to extend the signal in Fig. 4 and make it periodic, however, Fig. 5 doesn't seem to be such a valid extension.

I have stated that these signals have the same (identical) amplitude spectrum, which is quite different from the writer's quote above. Any time limited signal can be decomposed in a Fourier series, there is no need for periodicity. This is a good thing as any measured or recorded signal is time limited (even if it would last for a century). The Fourier coefficients can be obtained as described in the appendix of the article. However, with the Fourier synthesis it is possible to calculate the signal outside the original time interval. It shows that this is a repetition of the original signal. So if the original signal is recorded in the interval [0,T], the Fourier synthesis for the interval [T,2T] will result in the same signal as in the interval [0,T]. Or, in other words f(t+T) = f(t) and f(t+2T) = f(t), etc. It also holds for times < 0. This is caused by the periodicity of the (co)sines used in the synthesis. But if you want to only apply Fourier theory to signals from (-∞,∞), I hereby grant you permission to extend the signal of fig. 4 over this interval by repeating the signal ad infinitum. I think, however, it is not very interesting for the reader to see an infinite repetition of the same signal. Figure 5 is obtained as described in the appendix and it is certainly ***not*** *a periodic extension of fig. 4.*

To add insult to injury, even if it would be, a Fejér theorem corollary proves that if two functions f(x) and g(x) have the same Fourier coefficients, f≡g, therefore there are no two different functions having the same Fourier representation. As it is, the whole premise for Section 4 crumbles.

A function, as we encounter in audio, can be Fourier Transformed and this will lead to a one-to-one projection from time to frequency domains. This means that each function of time has only one equivalent in the frequency domain and vice versa. A major consequence is that any modification in the frequency domain will lead to a different function in time domain. The common misunderstanding is that leaving the amplitude spectrum unscathed will result in signals which are identical. This is, however, incorrect. It is also ***required*** *that the phase spectrum remains unscathed. Your comments highlight a common misunderstanding of this issue. As described in detail in the appendix, I have modified the phase spectrum of the signal of fig. 5 relative to the phase spectrum of fig. 4. Again, the details can be found in the appendix, but the results are rather severe and this clearly shows the point I want to make. A careful reading of the procedure described in the appendix, which can for instance be applied in MathLab, will make clear that keeping only the amplitude spectrum the same is no guarantee that the signals are the same in the time domain. A consequence of this is that a sound system with an incorrect phase response will modify the envelope of the reproduced sound and in my view, this is effecting the sound quality, which can e.g. be heard with drums and the attack of a grand piano. Another consequence is that the response of our hearing will also be different when the phase spectrum is modified. I have demonstrated this by converting the signals of fig. 4 and 5 into .wav files and playing these.*

Let's advance to Section 5. The author shows in Fig. 6 the response of a CD reconstruction filter to a complex signal. The author is not providing all the details of this experiment; is it the result of measurements of a certain reconstruction filter type, or is it just a theoretical analysis of the impulse response of the reconstruction filter? Whatever it is, this is an irrelevant result, since it does not reflect the reality of the CD signal processing. It is easy to determine the spectral content of the (periodic) signal in the upper half of Fig. 6. Given the A/D sampling frequency, an antialiasing filter is always required to filter out the input signal spectral components. Once that all the extra spectral components are removed, the D/A reconstruction is guaranteed to be a unique reconstruction of the input signal, as the Nyquist theorem applies. I would assume that the author confuses the A/D input antialiasing filter with the reconstruction filter at the D/A output? That would lead to the conclusion that the author believes the 44.1KHz CD sampling frequency is too small? Possible, but we are still far from proving that this is anywhere an audible limitation of the CD format.

To avoid too much detail at that stage, I showed that the filtering of the CD format leads to similar time smear as moving magnet pick-up cartridges, which can clearly be heard. This points, at least, to the necessity for further investigation. In my view, the sampling frequency of 44.1 kHz is indeed (far) to low because of the resulting time smear. See also below.

The author further claims “the definition of the Signal-to-Noise Ratio of a digital system is not really straightforward”. Well, there is no digital system specific SNR definition – it is the same as for analog systems that is P_{signal}/P_{noise}. What the author defines as “noise” appears to be the usual definition of quantization noise (an analog error signal summed with the signal before quantization), so let’s assume this is what the author means.

The fact that quantization noise is larger for low sampling frequencies is almost obvious and definitely well known; I don’t think there’s anything to prove here. Fig. 7 and 8 are illustrating this, and I would assume that the analog input signal is the same as in Fig. 6. However, I fail to see the relevance of this “simulation”, setting aside the fact that the signal in Fig. 6 is not related in any way, shape or form to the real “music”. Such signals could be maybe used for system identification (similar to a step input/response) but definitely not for estimating the “music” response. Otherwise, the sum of all confusions is in the author’s statement: “Note that this phenomena show up with discontinuous signals and not with “infinite” sine waves”. What are “discontinuous signals”? Assuming they are time limited signals like in Fig. 6, what is the relevance of this?

One example of a discontinuous signal in this context is the signal from a triangle. Dithering is always advocated as THE technique to avoid quantization noise and everybody shows the results for continuous sine waves. But by looking at shorter signals, one finds that this is untrue because the statistics do not work (yet). So the advantage of dithering is less than commonly stated.

Such signals do not have a direct Fourier representation (since they are not periodic) while their Fourier transformation (after extending the signal to +/-∞) has a spectrum that is truncated by the input antialiasing filter. The result is what it is, and the Nyquist theorem guarantees an exact reconstruction of the filtered input at the output. No good enough? Go for a higher sampling frequency (and a higher cutoff frequency for the anti-alias filter). But this has nothing to do with the CD sound quality, unless the author could somehow prove that signals as in Fig. 6 are anywhere relevant for music. Otherwise, I may suggest an even more striking example of “time domain distortions”, just use at the input a 50uSec square pulse and see what you get after 44.1KHz sampling.

A function can be Fourier transformed when it fulfils the requirements as the writer has listed earlier. Audio signals of finite duration fulfill these requirements. Which is good as else it would be impossible to calculate the spectrum of measured signals which are always time limited. See also my previous replies. See the appendix for the way to do it (which I bluntly, but happily, stole from standard textbooks on Fourier theory).

The discussion on high resolution formats is partly about the influence of frequencies above the sine wave hearing limit of 20 kHz. As the severe filtering, required with the use of 44.1 kHz sampling, both for the anti-aliasing as well as the reconstruction filter, will lead to severe time smear, this could be audible as the experiences with moving magnet and moving coil pick-up cartridges suggest. Improving the impulse response of our audio systems has lead to an improved sound quality, which also points in the same direction. In this way it might be possible that the actual presence of frequencies above 20 kHz is not the cause of the differences, but the time-smear, caused by the roll-off above 20 kHz.

I am not sure I understand Fig. 10 and Fig. 11, so I will not comment. Interesting enough, I don't necessary disagree with the author's conclusion so far: "12 bit system sounded so bad…noisy artifacts…" and neither do I disagree or agree to the statement "introduction of another 4 bit will certainly reduce the problem, but claiming that these effects will then be inaudible is, diplomatically put, questionable" simply because nothing was proved so far. The author formulates a conclusion without a shred of coherent proof.

Another blank statement is about a Dutch section of the AES in Eindhoven experiment, subjectively comparing the CD and SACD formats. Invoking the AES sound impressive, would it be too much to ask for a reference about this event? Further, well documented results published by AES ([1], [2]) seem to be in opposition (unless the author belongs to the ABX denial group, in which case the discussion would stop right away.

When a 12-bit system (with a theoretical SNR of 66 dB) sounds very bad (and the many witnesses of this demonstration have all confirmed this), it is, diplomatically put, very likely that the 24 dB additional SNR created by moving from 12 to 16 bit amplitude resolution will be insufficient to reduce the artifacts to inaudible levels. The 12 bit system is clearly inferior to a good analog system with a similar SNR, so what more proof does the writer want? I referred to a demonstration by Philips for the Dutch section of the AES, where the artifacts were still clearly audible with the CD version of a recording, but had disappeared with the SACD version of the same recording. The paper referenced by Mr. Popa himself [2], actually states that the"noise"in the CD reproduction enabled some listeners to distinguish the CD recordings from the hi-res recordings. So I do not understand this comment; it is rather obvious and reported in literature.

Now, where is this coming from, "These results show that the application of linear theory to digital systems is incorrect"? I have to admit I have no idea what the author infers here: which linear theory, how would this theory apply to digital systems, why would that application be incorrect, and who did that capital crime.

Again, it is essential to understand the difference between function and system. The comments of the writer in the first part of this letter make me fear that this basic understanding is lacking. I am therefore not surprised that the writer does not understand what I mean. Maybe my explanation will help him in this respect.

In the general presentations of digital systems linear theory is applied to calculate responses etc. As a digital system is inherently non-linear (to be or not to be, that's the question), this approach leads to incorrect results for non- continuous signals, albeit that the predictions for monotonous, continuous sine waves are pretty good. I had a discussion e.g. with Professor Vanderkooy about this because he had come to results which were in disagreement with mine and he had to admit that he had overlooked the problems with non-continuous signals. Music happens to be non-monotonous and non-continuous, for which I am very grateful.

It becomes progressively difficult to analyze the author's statements; at the end of Section 5. almost every phrase is either a blank statement or a statement without any shred of proof. Take this "the steep anti-aliasing filter results in time smear of the input signal which is worse than with MM pickup cartridges". Who said that?

I did and look at the figures in the paper: you can see it for yourself. But it is simple to add that the response of a MM cartridge is a 4th order low pass filter whereas CD anti-alias and reconstruction filters need to be of a significantly higher order to avoid aliasing. The Fourier inversion theorem states that steeper filters (at the same cut-off frequency) will have more time-smear (a longer impulse response).

Can we see some proof that this is a) true and b) audible? Or this: "the interaction of amplitude and time quantization leads to irregular modulated noise contributions, which are likely to give rise to loss of details", what exactly is the author talking about? Perhaps about the fact that quantization noise is non-linear and signal dependent and different calculations exist for different signal models? That's a well-established fact, but what would be the relevance here? And how exactly do we avoid "likely" and evaluate "loss of details"?

The further evaluation of what can be heard will take a lot more effort. I have tried to bring aspects to the discussion to help this evaluation. I do not have all the answers, but I have noticed a number of pitfalls which could invalidate the results of evaluations if these are not avoided. In this respect I would like to remind you that relatively recent evaluations of the "lossy" compression techniques concluded unambiguously that the bit rate reduction by the MP3 format was inaudible. Nowadays, we know that this is, diplomatically put, a bit beside the truth.

However, I cannot agree more to the author's introspection in the last paragraph, about "currently used 16bit/44.1KHz is sufficient to be inaudible": "In my view (or should I say "in my hearing"?) it is not". Full stop and it would be great to leave it here. I have absolutely no problems of assuming that the author is profoundly unhappy with the CD format quality. It's only the attempt to justify his dislike using flawed or pseudo-scientific arguments that I find annoying.

Unfortunately, as I noted in my replies above, I have to conclude that several of the "arguments" by the writer of this letter are flawed.

Section 6 repeats the same argumentation for the SACD format. The same comments as above apply – signals that have no relevance for music reproduction, misunderstanding of the Fourier transform and spectra, confusions between the input anti-aliasing filter and the output reconstruction filter, etc…

The bold statement in Section 6, (about "no sound reproduction in the world which can reproduce what our hearing can detect" is a well-known stance among hardcore subjectivists; they claim that the ear is some sort of super receiver, that beats (e.g.) the noise performance of the Arecibo radio astronomy dishes. No comments here, other than about those pesky scientific experiments, that the subjectivist camp are so deadly against, proving that our ear couldn't be a worse Fourier analyzer.

So far, I have never ever heard a sound reproduction system which presents the level of dynamics and details which I can distinguish during a live performance of e.g. a symphony orchestra. And that includes systems which are priced in the € 300 000,- to 400 000,- (roughly US$ 400 000,- to 500 000,-) range. I would welcome any reference to a system which would be able to do this, but until that day, I will stick to my statement that our hearing outhears any current sound reproduction system.

[1] http://www.bostonaudiosociety.org/bas_speaker/abx_testing2.htm

[2] "Audibility of a CD-Standard A/DA/A Loop Inserted into High-Resolution Audio Playback JAES Volume 55 Issue 9 pp. 775-779; September 2007 Authors: Meyer, E. Brad; Moran, David R.

This last referenced paper is one of the most criticized papers in the AES Technical Committee on High Resolution Audio, as well as in our Dutch section. This is due to a number of inconsistencies in the used procedure.