Apply the sampling theorem correctly!

The influence of aliasing filters as the "source of all evil" has become a subject of discussion more and more often. Already in my first article many years ago, I discussed this and included many beautifully illustrated diagrams; admittedly only one aspect among many. Afterwards, I went on to discuss something completely different. Since then, after having many new experiences and reflections based on them, I am now inclined to reduce the entire topic of "choosing the correct sampling rate for music recording" to filters, and therefore I would like to discuss this issue again. Based on current information, I would reduce everything down to one sentence:

"Apply the sampling theorem correctly!"

In principle, everyone knows the saying: "the sampling rate has to be double the highest frequency to be transmitted". I think no one will object to this. But as a rule, this principle is usually applied in exactly the opposite way: the bandwidth is limited to half of the desired sampling rate. At this point, the limits of human hearing are usually brought into the argument. However, this is a false comparison. The only crucial comparative figure is – just as the sampling theorem states – the bandwidth of the useful signal. Because the bandwidth is reduced with the help of technical equipment (aliasing filter), the signal is changed in the useable band as well. Initially, it was believed that using such wonderfully symmetrical digital filters so stable in their phasing was being on the safe side. Today the term ringing is known to all of us. These filters generate temporal smears and new spectral components which are not part of the original signal and unfortunately within the limits of human perception.

So the decisive question is not whether people have any use for frequencies above 20 kHz. These frequency components are simply part of the music signal to be digitalised, because instruments produce them, microphones record them and analog amplifier technology transmit them. If the sampling theorem is not observed correctly, but only carried through the "back door", then the filters employed will unavoidably disturb the signal. If the sampling theorem is applied correctly – so if the sampling rate is really at least double the highest frequency contained in the useful signal – then these errors do not occur and the music signal will contain no artefacts. Therefore – and maybe really only therefore – sampling rates which are higher than 48 kHz sound better than those below this rate. Already at 96 kHz, filters carry significantly fewer artefacts and at 192 kHz there are quasi non-existent, because there are no more to be found in the frequency spectrum of music which need to be filtered out. Only starting at a sampling rate of 192 kHz is the sampling theorem fully and correctly observed.

Any type of argumentation made, such as at the time the CD was invented and again today, "a sampling frequency of 44.1 kHz is satisfactory in order to store everything the human ear can hear in digital form", is wrong. In contrast, the following would be correct: "A sampling frequency of 44.1 kHz is satisfactory in order to store everything contained in the useful signal in digital form." But if this is true, it is only true for exceptional cases in music. In my view, understanding this subtle difference is the heart of the matter!

The bandwidth of the useful signal as it leaves the microphone sets the reference size on which everything is based, if the transmission is to succeed as being as true as possible. If the recording is not based on this value, then the quality is automatically lower (see "A Mathematical Theory of Communication" of 1948 by Claude Elwood Shannon, pgs. 47-48).

DSD

The central construction error of the CD format – the negative influences of the filters – were apparently already recognised in Sony's research department in the mid-1990s. Back then, the delta sigma theory was widely applied to audio converters. The actual converter stages from and to analog work here with a bitstream on the digital side. Conversion stages convert these bitstreams into PCM or generate them from PCM. Simply leaving out these so-called decimators or interpolation filters and to transport the bitstream without these intermediary stages was the basic idea behind DSD and SACD. The fact that the aforementioned errors could be avoided without these filter stages is the essential advantage of DSD. When compared to the status of technology in the 1990s, the format could only mean an improvement (although I personally never felt this way when comparing the sound).

However, there are two considerable disadvantages: because of the principles involved, the audio signal in the DSD format contains relatively high noise levels in the high frequency range starting just below 20 kHz. However, DSD for the most part cannot be edited in the studio. Either it has to be edited in analog or converted to PCM. The way most often chosen is the latter, that is, if the entire project is not already produced in PCM anyway. There are probably only very few genuine DSD productions.

Technology advanced, of course and in today's view, this 20 year old idea has long been made obsolete. The delta-sigma theory has remained, but it has gone through an enormous develop process. The leading chip manufacturers mostly work with internal ADC and DAC chips with 12.288 MHz at 6 bit instead of 2.822 MHz at 1 bit, which was standard for SACD at the beginning of the 1990s, thus giving rise to the DSD format for SACDs. The logic at the time, to simply leave the decimators and interpolation filters out of the AD and DA converters and store the resulting format instead, has been superseded by technological development. The step from 1 bit to multibit was a decisive leap. If you do some research on the theory behind DSD, you will encounter the problem of idle tones – which is possibly the explanation for the aforementioned dissatisfaction with the sound. The problem is completely avoided with multibit.

On the other hand, negative interference of filters in PCM with current technology is easily avoided, since the delta-sigma bitstream is converted to PCM with a high enough sampling rate; high enough in the sense of observing the sampling theorem. The problems which DSD was supposed to solve in the mid-1990s were solved in the contemporary use of PCM. When using 24/192, there are no filter artefacts and the bandwidth is definitely considerably larger than that of DSD. The impulse solutions which are readily shown regarding this are misleading and have the DSD be more precise than it actually is for a typical music signal. Because as a rule, high frequencies in a music signal always have a comparatively low volume level. Therefore, analog tape machines and records work with the corresponding equalisations. **A transmission system for audio has to be able to process high frequencies as precisely as possible at low volume levels.** Even the inventors of the CD knew that, which is reflected in the emphasis option. The inventors of the SACD, however, apparently totally forgot what type of signal they were working with. Only the impulses which are so readily shown (Dirac delta function), are perfect for DSD. They include all frequencies at the same volume. Such in impulse is transmitted well per DSD with a volume level just under full modulation, because all spectral components are below the level of high frequency noise. Unfortunately, this has nothing to do with the reality of music transmission, because in music, the high frequency components disappear into noise. DSD actually has the opposite characteristics than what is necessary for music; this is why ancient record and tape recordings were optimised for this already way back when, for example. DSD can only transmit high frequencies at high volume levels. But music does not contain high frequencies at high volume levels. There are thus more realistic statements from tests with square wave signals – here, the amplitude of the n-th harmonic decreases with 1/n – and hardly surprising, it can be clearly recognised that PCM 24/192 is superior to DSD64. At best, DSD128 is interesting, because the noise begins about as high up as the bandwidth of 24/192 PCM ends. But another essential problem remains in spite of this: DSD cannot be re-edited in the studio. The digital conversion to PCM and back is anything but trivial and the resulting losses again call any possible uses into question. The same applies to the alternative production path via analog stages and the additional conversion processes which become necessary because of them. In contrast, the admittedly positive aspects of DSD regarding filter artefacts can also be achieved in PCM formats if the sampling rates are high enough. These formats are easily edited in the studio at a very high level of quality.

If the idea behind DSD is an up-to-date approach, then these 12.288MHz/6bit signals of modern converter chips have to be stored. It is doubtful that this immense effort compared to 24/192 PCM – which is well-known to be directly re-editable – is really sensible. On the other hand, the distance between optimally implemented 24/192 PCM and outstanding analog is already too small, if it exists at all. The renaissance of DSD is thus more to be seen in the context of the poorly understood design errors of first generation digital formats. The defects can be heard. But unfortunately, the wrong or at least less efficient solution is being chosen.

Ralf Koschnicke / ACOUSENCE