

Das Abtasttheorem richtig anwenden!

Der Einfluss der Aliasingfilter als „Quell allen Übels“ wird ja in letzter Zeit immer öfter thematisiert. Schon in meinem ersten Artikel vor vielen Jahren hatte ich das auch angesprochen, inkl. vieler schöner illustrierender Diagramme; nur zugegeben als ein Aspekt unter vielen. In der Folge ging die Diskussion dann vornehmlich in eine andere Richtung. Inzwischen bin ich in der Summe vieler neuer Erfahrungen und der darauf gründenden Überlegungen jedoch durchaus geneigt, das ganze Thema „Wahl der für Musikübertragung richtigen Abtastrate“ im Prinzip auf die Filter zu reduzieren, und deshalb möchte ich mich doch nochmal zu diesem Thema äußern. Denn nach heutigem Stand würde ich eigentlich alles auf einem Satz reduzieren:

„Das Abtasttheorem richtig anwenden!“

Im Prinzip kennt ja jeder den Merksatz: „Die Abtastrate muss doppelt so hoch wie die höchste zu übertragende Frequenz sein.“ Dem widerspricht denke ich niemand. Nur wird der Satz doch in der Regel genau umgekehrt angewendet: Die Bandbreite wird auf die Hälfte der gewünschten Abtastrate begrenzt. An dieser Stelle wird dann mit dem menschlichen Hörbereich argumentiert. Das ist hier jedoch die falsche Vergleichsgröße. Die einzig entscheidende Vergleichsgröße ist – ganz wie das Abtasttheorem sagt – die Bandbreite des Nutzsignals. Denn wird die Bandbreite mit Hilfe von technischen Apparaturen reduziert (Aliasingfilter), bleibt das Signal auch im Nutzband nicht unverändert. Anfangs glaubte man sich mit den so wunderbar phasenstabilen, symmetrischen Digitalfiltern auf der sicheren Seite. Heute ist der Begriff Ringing hinlänglich bekannt. Diese Filter generieren nämlich zeitliche Verschmierungen und neue spektrale Komponenten, die im Originalsignal nicht enthalten sind und leider auch nicht außerhalb des menschlichen Wahrnehmungsbereichs liegen.

Entscheidend ist deshalb gar nicht die Frage, ob der Mensch irgendeinen Nutzen aus Frequenzen oberhalb 20kHz zieht. Diese Frequenzanteile sind im zu digitalisierenden Musiksignal einfach da, weil Instrumente sie erzeugen, Mikrofone sie aufnehmen und analoge Verstärkertechnik sie überträgt. Wird das Abtasttheorem nun nicht auf die korrekte Art und Weise, sondern nur durch „die Hintertür“ erfüllt, diskreditieren die dabei eingesetzten Filter das Signal unvermeidlich. Wird das Abtasttheorem richtig angewendet – wird also die Abtastrate tatsächlich mindestens doppelt so hoch wie die höchste im Nutzsignal enthaltene Frequenz gewählt – dann passieren diese Fehler nicht und es entstehen keine sich dem Musiksignal überlagernden Artefakte. Deshalb – und vielleicht wirklich nur deshalb – klingen höhere Abtastraten als 48kHz besser als die darunter. Schon bei 96kHz arbeiten die Filter mit deutlich weniger Artefakten und bei 192kHz existieren diese quasi nicht mehr, weil im Frequenzspektrum von Musik jetzt wirklich nichts mehr zu finden ist, was weggefiltert werden müsste. Erst ab 192kHz Abtastrate ist das Abtasttheorem in korrekter Auslegung vollständig erfüllt.

Falsch ist also jede Art von Argumentation, wie bei Erfindung der CD und auch gerade mal wieder gelesen: „Eine Samplingfrequenz von 44,1kHz ist ausreichend, um alles, was das menschliche Ohr hört, in digitalisierter Form abzuspeichern.“ Zulässig wäre hingegen allenfalls eine Formulierung wie „Eine Samplingfrequenz von 44,1kHz ist ausreichend, um alles, was im Nutzsignal enthalten ist, in digitalisierter Form abzuspeichern.“, wenn es denn stimmen würde, denn das ist bei Musik nur in Ausnahmefällen gegeben. Diesen feinen Unterschied zu begreifen ist aus meiner Sicht der Kern des Themas!

Die Bandbreite des Nutzsignals, wie es das Mikrophon verlässt, setzt die Referenzgröße nach der sich alles richten muss, soll die Übertragung möglichst ideal gelingen. Wird sich nicht danach gerichtet, wird automatisch ein niedrigeres Qualitätslevel gesetzt (siehe auch Claude Elwood Shannons „A Mathematical Theory of Communication“ von 1948, Seite 47/48).

DSD

Den zentralen Konstruktionsfehler beim CD-Format – die Negativeinflüsse der Filter – hatte man in der Entwicklungsabteilung bei Sony Mitte der 1990er Jahre offenbar auch erkannt. Des Weiteren hatte sich als Wandlertyp im Audibereich damals das Delta-Sigma-Prinzip weitgehend durchgesetzt. Die eigentlichen Wandlerstufen von und zu Analog arbeiten hier auf der digitalen Seite mit einem Bitstream. Umrechnungsstufen wandeln diesen Bitstream in PCM bzw. generieren ihn von PCM. Diese sogenannten Dezimationsfilter bzw. Interpolationsfilter einfach wegzulassen und den Bitstream ohne diese Zwischenstufen zu transportieren, war die Grundidee hinter DSD bzw. der SACD. Dass ohne diese Filterstufen die o.g. Fehler vermieden werden, ist der wesentliche Vorteil von DSD. Gemessen an dem Stand der Technik in den 1990er Jahren konnte das Format eigentlich nur eine Verbesserung bedeuten (Obwohl ich persönlich das akustisch nie so empfunden habe).

Es gab jedoch auch zwei entscheidende Nachteile: Prinzip bedingt enthält das Audiosignal im DSD-Format ab knapp 20kHz Rauschen im Hochtonbereich mit relativ hohem Pegel. Hauptsächlich jedoch lässt sich DSD studioteknisch nicht bearbeiten. Es muss entweder analog gearbeitet werden oder in PCM konvertiert. Letzteres ist der meist gewählte Weg, wenn nicht ohnehin komplett in PCM produziert wurde. Wirklich echte DSD-Produktionen dürfte es nur ganz wenige geben.

Die technische Entwicklung ging sodann weiter und aus heutiger Sicht ist diese 20 Jahre alte Idee längst überholt. Das Sigma-Delta-Prinzip ist zwar geblieben, es hat jedoch einen enormen Entwicklungsprozess durchgemacht. Die führenden Chiphersteller arbeiten heute intern in ADC- bzw. DAC-Chips meist mit 12,288MHz bei 6Bit statt mit 2,822MHz bei 1Bit, was bei Erfindung der SACD Anfang der 1990er Jahre Standard war und weshalb das DSD-Format für die SACD so geboren wurde. Die damalige Logik, man lasse einfach die Dezimations- und Interpolationsfilter in den AD- bzw. DA-Wandlern weg und speichere stattdessen dieses Format, wurde von der technischen Entwicklung überholt. Dabei ist der Schritt weg von 1Bit zu Multibit ein ganz entscheidender. Beschäftigt man sich mit der Theorie hinter DSD, begegnet einem das Problem der Idle-Tones – vielleicht die Erklärung für o.g. akustische Unzufriedenheit. Das Problem wird mit Multibit vollständig vermieden.

Auf der anderen Seite sind die Negativeinflüsse der Filter bei PCM mit aktueller Technik leicht vermeidbar, wird der Delta-Sigma-Bitstream nämlich auf PCM genügend hoher Abtastrate umgesetzt; genügend hoch im Sinne der Erfüllung des Abtasttheorems. Das was DSD Mitte der 1990er Jahre lösen sollte, stellt sich bei zeitgemäßem Einsatz von PCM nicht mehr als Problem. Bei 24/192 gibt es keine Filterartefakte und die Bandbreite ist auf jeden Fall deutlich größer als bei DSD. Die gerne gezeigten Impulsantworten sind diesbezüglich nämlich irreführend und lassen DSD präziser wirken als es für ein typisches Musiksignal wirklich ist. Denn grundsätzlich haben hohe Frequenzen in einem Musiksignal immer vergleichsweise niedrige Pegel. Deshalb arbeiten analoge Bandmaschine und Schallplatte ja auch mit entsprechenden Entzerrungen. **Ein Übertragungssystem für Audio muss hohe Frequenzen bei niedrigem Pegel möglichst exakt verarbeiten können.** Selbst den Erfindern der CD war das noch bewusst, was sich in der Emphasis-Option zeigt. Die Erfinder der SACD hatten aber anscheinend schon völlig ausgeblendet, mit welchem Signaltypus sie es zu tun haben. Lediglich die so gerne gezeigten Impulse (Dirac-Stoß) sind perfekt für DSD. Sie bestehen aus allen Frequenzen bei gleichem Pegel. Ein solcher Impuls wird per DSD bei einem Pegel knapp unter Vollaussteuerung somit auch sehr gut übertragen, weil alle spektralen Anteile über dem HF-Rauschen liegen. Mit der Realität der Musikübertragung hat das aber gar nichts zu tun, denn bei Musik verschwinden die hohen Frequenzanteile im Rauschen. DSD hat eigentlich genau das Gegenteil der Charakteristik, wie sie für Musik notwendig ist und worauf beispielsweise schon die uralten Schallplatten- und Bandaufzeichnungen damals optimiert wurden. DSD kann hohe Frequenzen nur bei hohen Pegeln übertragen. Musik enthält aber keine hohen Frequenzen bei hohen Pegeln. Realistischere Aussagen erhält man deshalb bei Tests mit Rechtecksignalen – dort sinkt die Amplitude der n-ten Oberschwingung mit $1/n$ – und man erkennt dann auch wenig überraschend sehr deutlich: PCM24/192 ist klar besser als DSD64. Interessant wird es allenfalls ab DSD128, weil das Rauschen dann ähnlich weit oben beginnt wie die Bandbreite bei 24/192 PCM endet. Aber ein anderes ganz wesentliches Problem bleibt trotzdem: DSD kann studioteknisch nicht weiterverarbeitet werden. Die digitalen Umrechnungsprozeduren zu PCM und zurück sind dann auch noch alles andere als trivial und die Verluste hierbei stellen den möglicherweise vorher bestehenden Nutzen wieder in Frage. Gleiches gilt für den alternativen Produktionsweg über analoge Stufen und die dadurch nötig werdenden zusätzlichen Wandlungsprozesse. Der zugegebene positive Aspekt bei DSD bezüglich Filterartefakten lässt sich hingegen auch in PCM-Formaten mit genügend hoher Abtastrate erreichen. Diese Formate lassen sich aber studioteknisch leicht auf sehr hohem Qualitätsniveau handhaben.

Soll die Idee hinter DSD zeitgemäß sein, müsste man also diese 12,288MHz/6Bit-Signale moderner Wandlerchips speichern. Ob dieser immense Aufwand im Vergleich zum 24/192 PCM – das ja bekanntermaßen direkt weiterverarbeitet werden kann – dann aber wirklich sinnvoll erscheint, ist zu bezweifeln. Dafür ist der Abstand zwischen optimal umgesetztem 24/192 PCM und einem noch so hervorragendem Analog bereits zu klein, wenn überhaupt noch vorhanden. Die Renaissance von DSD ist eher im Kontext des nicht wirklich verstandenen Konstruktionsfehlers bei digitalen Formaten erster Generation zu sehen. Der Mangel wird zwar akustisch erkannt. Es wird aber der falsche oder zumindest ineffektivere Lösungsweg gewählt.