

» Das kleine Einmaleins der „Digitalen Audiologie“

WISSEN SCHAFFT AUDIO



Wir sind auf Schneller-Höher-Weiter konditioniert: Egal ob es sich um Gigahertz, Megapixel oder Samplingrate handelt, wir vergleichen Geräte häufig nach dem Motto „viel hilft viel“. Doch so lange das Maß aller Dinge immer noch unsere Augen und Ohren sind, sind solche Daten erschreckend untauglich, um die effektive Qualität eines Systems zu ermessen. In der Praxis zählen oft ganz andere Dinge, die sich nur nicht in einer griffigen Zahl ausdrücken lassen. Und das Beste: Dieses Wissen kostet nichts.

von Roman Beilharz

» Wer den eigenen Vorstellungen und Möglichkeiten entsprechend optimale Aufnahmen machen möchte, kommt früher oder später nicht mehr drum herum, sich eingehender mit dem Thema „Digital-Audio“ zu befassen. Aliasing, Shannon'sches Theorem, Dithering, Oversampling – fast jeder hat diese Worte schon mal irgendwo gehört oder gelesen. Die exakte Bedeutung dieser Begriffe ist aber selbst professionell arbeitenden Tonschaffenden nicht immer

wirklich klar. Dabei ist es nicht einmal nötig, deren wissenschaftliche Grundlagen erschöpfend zu studieren. Was zählt, ist das kleine Einmaleins der „Digital-Audiologie“. Um dieses korrekt zu vermitteln, müssen wir ganz von vorne anfangen; komplizierter wird es dann von ganz alleine. Zunächst ist es unerlässlich, zwei Dinge strikt auseinander zu halten: Die Samplingrate und die Wortbreite (Bit-Tiefe). Die Samplingrate bestimmt, wie oft pro Sekunde die Spannung einer analo-

gen Eingangsschwingung in einem Datenwort – dem so genannten Sample – gespeichert wird. Die Wortbreite bestimmt dagegen, wie präzise diese analoge Eingangsspannung durch den Wert eines Samples angenähert werden kann. Dieser Wert wird binär codiert – also in Bits mit dem Zustand 1 oder 0 umgeformt – und in 8er-Paketen gespeichert; letztere nennt man Bytes. Ein 16-Bit-Sample für einen Kanal (Mono) benötigt also 2 Byte Speicherplatz. Bei einer Samplingrate von 44,1 kHz (44.100 Mal pro Sekunde abgetastet) in 16 Bit ergibt sich demnach pro Sekunde ein Speicherbedarf von 88200 Bytes für ein Mono-Signal (44.100 x 2 Byte) und 176.400 Bytes für ein Stereo-Signal (44.100 x 2 Byte x 2 Kanäle). Heute werden fast ausschließlich Wortbreiten von 16 oder 24 Bit verwendet. Ein 16-Bit-Sample kann 65.536 verschiedene Spannungswerte darstellen, ein 24-Bit-Sample immerhin 16.777.216. Da jede analoge Welle zwischen positiven und negativen Signalspannungen hin und her schwingt, stehen für jede Halbwelle – damit meint man den Teil der Welle ober- oder unterhalb von

0 Volt – die Hälfte aller möglichen Digitalwerte zur Verfügung. Liegt die analoge Eingangsspannung des Ursprungssignals irgendwo zwischen zwei möglichen Daten-Werten, so findet zwangsläufig eine Rundung statt (Quantisierung). Das beschriebene Digitalisierungs-Verfahren nennt sich daher auch Spannungsquantisierung oder auf neu-deutsch Pulse Code Modulation – kurz PCM. So weit, so einfach. Doch die vielgestaltigen Probleme dieses Verfahrens stecken im Detail, sodass wir uns beide Ebenen – die zeitliche Auflösung (Samplingrate) und die dynamische Präzision (Wortbreite) – erstmal getrennt ansehen sollten.

Die Samplingrate

Im Studiobereich werden aktuell hauptsächlich Raten von 44,1/88,2 kHz (CD) sowie 48/96 kHz (DVD/TV/Film) verwendet. In vielen Harddisk-Recording-Programmen lassen sich Audiodateien verschiedener Samplingraten nicht gleichzeitig ohne falsche Tonhöhen-Wiedergabe verwenden, denn Soundkarte, Treiber und Software geben dem System in der Regel eine feste Samplingrate vor. So klingen z. B. 48 kHz-Dateien in einem 44,1 kHz-System zu langsam und tiefer. Doch selbst wenn ihre Recording-Software Audio-Dateien mit unterschiedlichen Samplingraten ohne zu murren korrekt abspielt, so geht dieser Komfort zu Lasten der Audioqualität. Um dies zu vermeiden, empfiehlt es sich, vor Beginn einer Produktion eine einheitliche Samplingrate für alle Arbeitsschritte bis zum finalen Medium festzulegen. Die benutzte Taktfrequenz sollte der Standard-Samplingrate dieses Mediums entsprechen – oder einem gradzahligen Vielfachen davon. Damit minimiert man Rundungsfehler beim abschließenden Herunterrechnen der Arbeits-Samplingrate (z. B. 88,2 kHz) auf die Samplingrate des Masters (z. B. 44,1 kHz/CD). Dieser Vorgang nennt sich Downsampling. Kombiniert man Dateien mit einfacher Samplingrate (44,1/48 kHz) im selben Projekt mit Dateien doppelter Samplingrate (88,2/96 kHz), so sollten die Dateien mit einfacher Rate im Vorfeld hoch gerechnet werden und die doppelte Rate als System-Takt gewählt werden. Dieser Vorgang nennt sich Upsampling und lässt sich mit jedem halbwegs brauchbaren Wave-Editor durchführen. Beim Upsampling durch exaktes Verdoppeln bzw. Vervielfachen der Samplingrate bleibt das Audiosignal zu 100 Prozent unverfälscht. „Krumme“ Up- oder Downsampling-Vorgänge – z. B. von 44,1 kHz auf 96 kHz oder umgekehrt – empfehlen sich nicht, denn hier sind Rundungsfehler unvermeidlich. Nur sehr

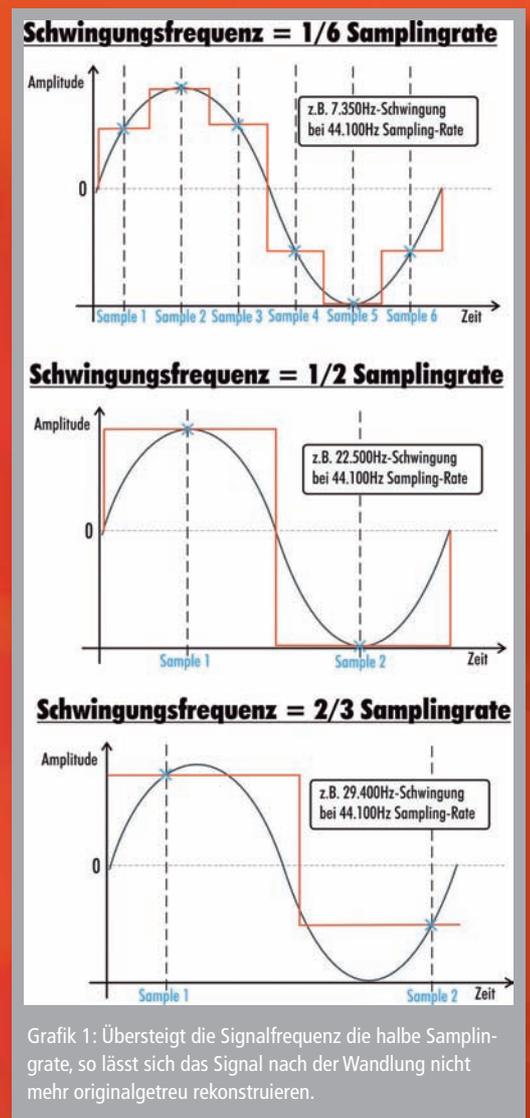
aufwändige Resampling-Algorithmen in Profi-Programmen erzielen auch hier gute Ergebnisse.

Nyquist-Shannon-Theorem

Um zu verstehen, welche Rolle die Samplingrate allgemein spielt, befassen wir uns mit Grafik 1. In der obersten Kurve sehen wir, dass sich unsere Schwingung gut durch 6 Samples repräsentieren lässt, da die Samplingrate sechs Mal so hoch ist, wie die Frequenz der Eingangsschwingung. Mit nur zwei Samples pro Schwingungswelle (Mitte) lässt sich die ursprüngliche Schwingung gerade noch nachbilden. Egal welcher Art die Eingangsschwingung im Original war – der DA-Wandler macht aus ihr eine Sinusschwingung, aber sie ist immerhin noch darstellbar. Ganz aus ist es, wenn weniger als zwei Samples pro Schwingung zur Verfügung stehen (unten). Das resultierende Digitalsignal weicht hier so erheblich von der Eingangsschwingung ab, dass es unbrauchbar wird. Anhand dieser Beispiele lässt sich leicht nachvollziehen, dass die Samplingrate mindestens doppelt so hoch sein muss wie die höchste Frequenz des Eingangssignals, damit ein brauchbarer Datenstrom entsteht: nur dann haben wir mindestens zwei Samples für jede vorkommende Schwingung zur Verfügung. Und das ist es dann auch schon, das geheimnisvolle Nyquist-Shannon'sche Abtasttheorem. Bei 44.100 Hz Samplingrate können demnach Frequenzen von bis zu 22.500 Hz wiedergegeben werden – theoretisch. Wenn man bedenkt, dass ein durchschnittlicher Erwachsener Frequenzen bis ca. 17.000 Hz wahrnimmt, dann erscheint diese Obergrenze mehr als ausreichend. Dummerweise enthalten analoge Audiosignale regelmäßig Frequenzanteile, die nicht nur weit über der Hörgrenze, sondern auch weit über der kritischen Schwelle der halben Samplingrate liegen (Nyquist-Frequenz). Ein Crash-Becken erzeugt z.B. Frequenzen bis weit über 80 kHz! Finden diese hohen Frequenzen ungefiltert Eingang in den Digitalisierungsprozess (auch Quantisierung genannt) so entstehen Signalverfälschungen, die sich später nicht mehr eliminieren lassen.

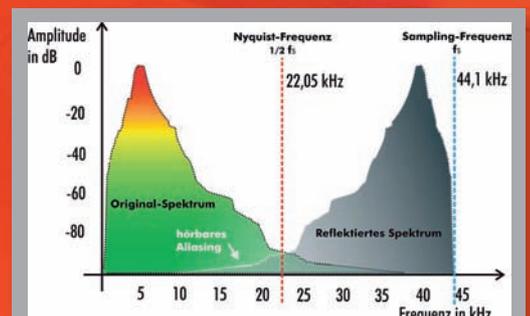
Aliasing

Die entstehenden Verfälschungen entstehen dadurch, dass nach einer ungefilterten Analog-Digital-Wandlung das Spektrum des Eingangssignals stets an der Frequenz, die der Samplingrate entspricht (fs), gespiegelt erscheint. Die Grafik 2 macht diesen eigenartigen Prozess klarer: Wir sehen deutlich, dass die Frequenzen oberhalb der halben



Grafik 1: Übersteigt die Signalfrequenz die halbe Samplingrate, so lässt sich das Signal nach der Wandlung nicht mehr originalgetreu rekonstruieren.

Samplingrate (1/2 fs) um 180 Grad gedreht in den hörbaren Bereich hinein „reflektiert“ wurden. Diese Aliase der Frequenzen des Originalsignals über 1/2 fs sind selbst bei geringen Obertonanteilen deutlich als Störgeräusche hörbar. Dieses Phänomen ist als Aliasing bekannt und eines der Hauptproble-



Grafik 2: Beim Aliasing werden durch „Spiegelung“ an der Sampling-Frequenz Frequenzen oberhalb der halben Samplingrate in den hörbaren Bereich reflektiert.

me bei der AD-Wandlung. Um dieses Problem in den Griff zu bekommen, müssen wir also die Frequenzen oberhalb der halben Samplingrate vor der Digitalisierung möglichst vollständig loswerden. Dazu könnte man nun dem Wandler einen Filter vorschalten, der alle Frequenzen unterhalb von 20 kHz unbehelligt ließe und bis zu 20,05 kHz maximal „zumachen“ würde (s. Grafik 3/Mitte). Solch ein ideales Anti-Aliasing-Filter lässt sich jedoch analog – es muss ja schließlich noch vor der AD-Wandlung eingreifen – nicht konstruieren. Denn mit einer solch extremen Flankensteilheit würde die Phasen- und Frequenztreue des Eingangssignals stark in Mitleidenschaft gezogen. Die so genannte Welligkeit analoger Filter würde besonders den Höhenbereich störend verändern (s. Grafik 3/unten). Abgesehen davon sind solche Filter in der Herstellung aufwändig und teuer.

Wir müssen uns also etwas Praktikableres ausdenken. Wenn es uns gelänge, das Aliasing-Spektrum aus Grafik 2 weiter nach oben (in der Grafik: rechts) zu verlegen, dann würde kein signifikanter Signalanteil der „Reflektion“ mehr in den hörbaren Bereich ragen. Wie aber kommen wir da hin? – Richtig: Die Samplingfrequenz f_s muss erhöht werden. Die Grafik 4 zeigt, dass sich unser Spektrum bei einer Verdoppelung der Samplingrate auf 88,2 kHz nicht mehr mit dessen Spiegelung überlappt und so auch ohne Filter keine sichtbaren Aliase mehr in den hörbaren Raum reflektieren.

Oversampling

Im wirklichen Leben reichen allerdings 88,2 kHz allerdings nicht aus, um Aliasing völlig zu verhindern – wir erinnern uns an das Crash-Becken mit Signalanteilen weit über

80kHz. Gängige AD-Wandler tasten das Eingangssignal daher mit mindestens dem 64-fachen ihrer Ausgaberate ab, das sind bei 44,1 kHz immerhin knapp 3 MHz (64-faches Oversampling). Die kritische Schwelle der halben Samplerate liegt nunmehr so hoch, dass selbst die höchsten Frequenzanteile des Analogsignals kein Aliasing mehr im Hörbereich verursachen. Der Vorteil ist nun, dass man den nach wie vor benötigten Filter zur Absenkung der „Spiegel-Spektren“ problemlos in den digitalen Bereich verlagern kann. Ein digitales Filter ist relativ billig zu realisieren und lässt sich problemlos in den Wandlerchip integrieren. Um die gewünschte Ausgabefrequenz von z.B. 44,1 kHz zu erhalten, muss die Samplerate des digitalen Signalstroms natürlich anschließend wieder reduziert werden. Oversampling hilft nicht nur, eine praxistaugliche Anti-Aliasing-Strategie zu entwickeln, sondern auch um eine höhere Signaltrou zu erzielen. Beim Sampling werden sehr schnelle Änderungen in der analogen Eingangsspannung zwischen zwei Samplepositionen nicht erfasst. Diese Ungenauigkeit geht bei hohem Oversampling gegen Null, da diese „Lücke“ extrem kurz wird. Zwar verlieren wir einen Teil der Signaltrou wieder beim Downsampling, jedoch ist ein Durchschnittswert aus z. B. 64 Werten an 64 Positionen immer noch genauer als ein einzelner Wert an einer einzelnen Position.

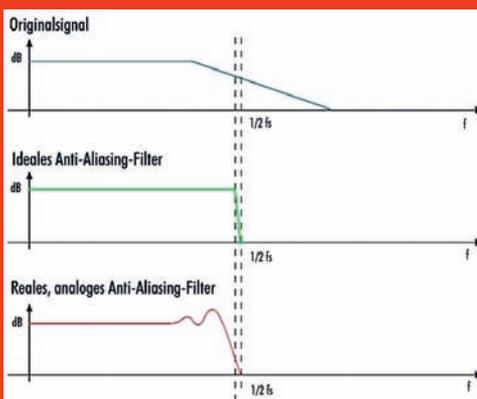
Intermodulations-Verzerrungen

Es könnte alles so schön sein, wenn das Nyquist-Shannon-Theorem nicht stillschweigend voraussetzen würde, dass der Abtastvorgang stets bei einem Nulldurchgang im Eingangssignal beginnt. Die obere Halbwelle einer 22,05 kHz-Sinus-Schwingung kann dann in der Tat wie in Grafik 1 (Mitte) durch

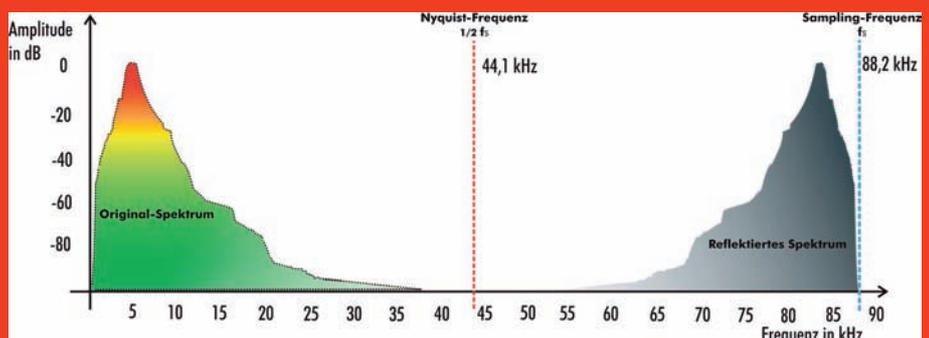
einen positiven, die untere Halbwelle durch einen negativen Sample-Wert beschrieben werden. In der Realität ist jedoch die Frequenz des Eingangs-Signals in keiner Weise mit der Digital-Clock des Wandler-Systems synchronisierbar – eine Tatsache mit dramatischen Folgen. Selbst wenn wir bei unserem relativ simplen Beispiel-Signal von 22,05 kHz Sinus bleiben, scheitert der praktische Versuch, dieses am Ausgang zu rekonstruieren in rund 50% aller Fälle. Denn wenn die Abtastperiode zufällig erst bei der Hälfte der oberen Halbwelle beginnt und damit zwangsläufig exakt bei der Hälfte der unteren Halbwelle endet, ist der resultierende Abtastwert gleich Null: die Schwingung geht vollständig über die Wupper (siehe Grafik 5C). Noch schlimmer wird es mit Musiksignalen, denn deren Frequenz und Phasenlage verhalten sich in Bezug auf die Abtastrate völlig chaotisch. Selbst bei einem einigermaßen konstant schwingenden Signal mit hohen, im Verhältnis zur Abtastrate ungradzahligen Frequenzanteilen, ändert sich die Phasenlage der beschreibenden Samples mit jeder Welle. Das Phänomen ist als Intermodulations-Effekt bekannt, die daraus resultierenden, mit zunehmender Frequenz immer stärker werdenden unharmonischen Verzerrungen nennt man auch Intermodulations-Verzerrungen.

Impulstreue

Wenn man ein Digital-Audio-System bei 44,1 kHz Samplingrate mit einer Rechteckschwingung füttert und sich ansieht, was von dieser nach der DA-Wandlung noch übrig ist, stößt man auf eine weitere unvorteilhafte Eigenschaft von Digital-Audio: dessen mangelhafte Impulstreue. Letztere wird von der Fähigkeit bestimmt, Ein- und Ausschwingvorgän-



Grafik 3: Ein ideales Anti-Aliasing-Filter lässt sich analog nicht preiswert konstruieren, denn mit der Flankensteilheit eines Filters nimmt auch dessen Welligkeit zu.



Grafik 4: Bei einer ausreichenden Vervielfachung der Samplingfrequenz gelangen keine Signalanteile des Alias-Spektrums mehr in den hörbaren Bereich.

ge präzise darzustellen. Auch analoge Systeme glänzen hier nicht immer: Lahme, unterdimensionierte Netzteile, minderwertige Bauteile und mangelhafte Kabelwege sind hier meist die Schuldigen. Je schneller ein Übertragungssystem in der Lage ist, Spannungswerte im Audio-Signal an- oder absteigen zu lassen, desto knackiger und „trockener“ klingen Bass, Gitarre, Klavier und Schlaginstrumente. Die Impulstreue ist anhand der Flankensteilheit eines Rechtecksignals messbar. Die traurige Wahrheit ist, dass ein Digitalsystem mit 44,1 kHz Abtastrate schon ein Rechtecksignal von gut 7 kHz aufwärts nicht mehr originalgetreu überträgt. Dieses kommt bis zur Unkenntlichkeit geglättet als Sinus-Schwingung wieder „hinten raus“. Die Grafik 6 macht dies deutlich: Die gerade mal 6 zur Verfügung stehenden Samples beschreiben die Kanten des Rechtecksignals nicht mehr ausreichend. Rechtecksignale mit höheren Frequenzen werden natürlich erst recht „weichgespült“. Damit arbeiten Digital-Systeme bei Standard-Abtastraten noch nicht einmal bis zur Hälfte des hörbaren Spektrums mit einer messtechnisch gesehen akzeptablen Impulstreue. Die Interpolations-Algorithmen der Wandler geben den

Schwingungs-Flanken den Rest – schließlich sollen diese ja Treppenartefakte verhindern und machen immer schön „einen auf Sinus“.

Jitter

Was machen wir denn nun, um die Impulstreue zu erhöhen und Intermodulations-Verzerrungen zu minimieren? Beide Phänomene gründen schließlich auf einer mangelhaften zeitlichen Auflösung für hohe Frequenzen –

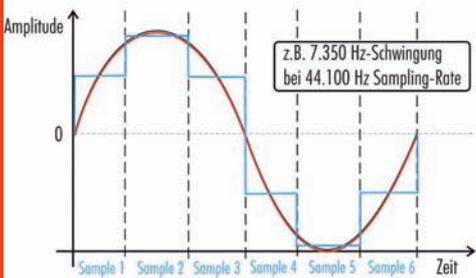
Eine Clock mit hohem Jitter verewigt sich bei der Aufnahme irreversibel im Sample-Datenstrom.

zumindest bei einfachen Samplingraten (44,1/48 kHz). Klarer Fall: Wir müssen die Arbeits-Samplingrate so hoch wie möglich wählen, dann wird alles gut. Theoretisch. Dummerweise klingen in der Praxis jedoch günstige AD-Wandler bei 44,1/48 kHz oft eindeutig besser als bei 192 kHz. Was ist da los? Dies liegt zum einen daran, dass die Filter-Strategien etlicher Wandler-Chips offenbar viel sorgfältiger für einfache Samplingraten abgestimmt wurden, als für vierfache, da letztere bei Homerecordern nur selten wirklich zum Einsatz kommen. Zum zweiten liegt

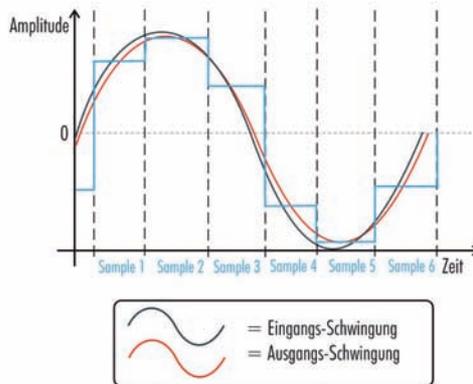
dies daran, dass Digital-Audio ein weiteres, häufig unterschätztes Übel mit sich bringt: Jitter. Dieser Begriff steht für die internen Schwankungen der digitalen Word-Clock, also des Taktgebers der Samplingrate selbst. Nur sehr teure und hochwertige Clock-Generatoren haben sehr geringe Takt-Schwankungen. Je höher nun die Samplingrate ist, desto schwieriger und aufwändiger wird es für die Entwickler, sehr gute Jitter-Werte (=geringe Takt-Schwankungen) zu erreichen. Daher ist es bei günstigen Audio-Interfaces meist so, dass sie zwar bei einfachen Samplingraten noch ganz ordentliche Jitter-Werte hinkriegen, doch bei doppelter oder gar vierfacher Samplingrate so „herumzap-peln“, dass die Jitter-Artefakte im Klang die theoretischen Vorteile der höheren zeitlichen Auflösung zunichte machen. Es gibt verschiedene Arten und Quellen von Jitter, auf die wir hier nicht in epischer Breite eingehen können. Hier nur so viel: Sind diese Schwankungen rein zufälliger Natur, verursachen sie ein amplitudenmoduliertes Rauschsignal, welches den Dynamikumfang des Systems verschlechtert. Dies kann man verschmerzen, bestehen die letzten 4 Bit eines 24-Bit-Signals ohnehin meist nur aus Rauschen. Schlimmer sind signalabhängige Schwankungen durch mangelhafte Kabel,

Schwingungsfrequenz = 1/6 Samplingrate

A) Abtastrate und Schwingungsfrequenz phasengleich

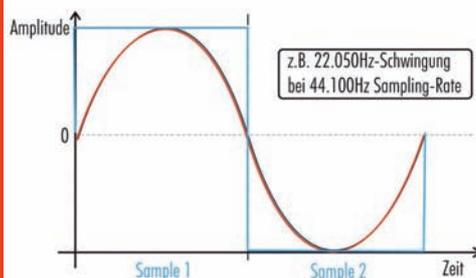


B) Abtastrate und Schwingungsfrequenz phasenverschoben

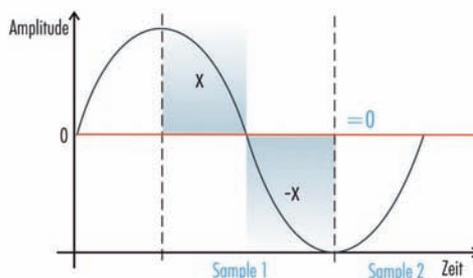


Schwingungsfrequenz = 1/2 Samplingrate

C) Abtastrate und Schwingungsfrequenz phasengleich



D) Abtastrate und Schwingungsfrequenz um 90° phasenverschoben



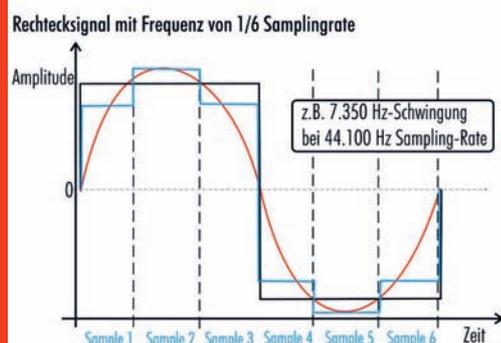
Grafik 5A: Bei einer Frequenz von 1/6 der Sampling-Rate lässt sich eine Sinus-Schwingung am Eingang (schwarz) bei phasengleichen Abtast-Perioden (blau) am Ausgang des DA-Wandlers perfekt rekonstruieren.

Grafik 5B: Dies gelingt selbst dann noch recht gut, wenn die Abtastfrequenz zur Frequenz der Eingangs-Schwingung phasenverschoben ist. Dies ist in der Praxis häufig der Fall, da sich die Phasenlage von Musiksignalen dauernd ändert.

Grafik 5C: Bei einer Schwingungsfrequenz von der halben Samplingrate wird es eng: Nur bei phasengleicher Abtastung lässt sich die Eingangs-Schwingung noch brauchbar am Ausgang wieder herstellen.

Grafik 5D: Ist die Abtastung derselben Sinus-Schwingung um 90° verschoben, heben sich die positiven und negativen Amplitudenwerte zu Null auf: Die Eingangs-Schwingung geht komplett verloren.

Impulstreue in Digital-Audio-Systemen



Grafik 6: Ein Digital-System mit 44,1 kHz Sampling-Rate „schleift“ ein Rechtecksignal ab 7 kHz aufwärts zu einer Sinus-Schwingung rund.



Erst ab der gehobenen Mittelklasse darf man bei Audio-Interfaces mit einer jitterarmen Clock rechnen.

schlechte Netzteile und minderwertige Wandler-Elektronik. Dann entstehen unharmonische Störsignale, welche gleichzeitig viel tiefer und viel höher als die anregende Signalfrequenz schwingen und daher kaum von dieser maskiert werden. Je lauter und hochfrequenter das Eingangssignal, desto größer werden die Abstände von Signalfrequenz zu Störsignal. Durch die hohe Dynamikauflösung aktueller Wandlersysteme werden solche Artefakte noch nicht mal gnädig vom Grundrauschen überdeckt. Im Hörtest macht sich Jitter als Klingeln in den Höhen, Rauheit im unteren Mittenband bzw. Schwammigkeit in den Bässen bemerkbar. Die Räumlichkeit und Differenziertheit leidet; das Klangbild tendiert in Richtung „verwaschen“ und „künstlich“.

Einmal Jitter, immer Jitter

Für die Praxis ist entscheidend, dass hohe Jitter-Werte besonders während der AD-Wandlung kritisch sind. Denn dann stimmt die zeitliche Position der Samples nicht genau genug mit der Position derselben bei der späteren DA-Wandlung überein und lässt sich auch nicht „naturgetreu“ wieder herstellen. Durch das ungleichmäßige Abtasten des analogen Eingangssignals kommt es zu Sampledaten, die sich nur dann zeitlich dem Originalsignal entsprechend korrekt wiedergeben lassen würden, wenn die DA-Wandlung mit einer exakt identischen Ungenauigkeit im Taktsignal stattfinden würde – was real nicht vorkommt. Anders ausgedrückt: Eine „verjitterte“ Aufnahme lässt sich mit keinem Plug-in der Welt nachträglich geradebiegen. Was jedoch funktioniert ist die

„Säuberung“ einer minderwertigen Clock durch einen hochwertigen, externen Taktgeber. Allerdings ist dieses Konzept in der Regel selbst bei der Verwendung von teuren Clock-Generatoren die schlechtere Wahl, als gleich einen Wandler mit hochwertiger interner Clock zu nutzen. Dieser kann die Vorteile kurzer interner, exakt vom Hersteller bestimmbarer Signalwege besser nutzen. Denn selbst bei der Verwendung von Qualitäts-Kabeln ist schwer vorauszusagen, wie sauber der jeweilige AD-Wandler sich auf die externe Word-Clock synchronisieren kann. Für die AD-Wandlung sollte daher stets das beste verfügbare Wandlersystem verwendet werden.

Bei der DA-Wandlung sorgen geringe Jitter-Werte zwar ebenfalls für ein präziseres, neutraleres Hörerlebnis, doch der Hörer wird

die Aufnahme vermutlich ohnehin mit einer „Gelee-Clock“ aus dem MP3- oder CD-Player abhören. Insofern ist ein DA-Wandler mit nur durchschnittlichen Jitter-Werten etwas, mit dem man beim Homerecording gut leben kann – solange man die schöne Gewissheit hat, dass die Aufnahmen selber relativ jitterfrei sind und auch auf hochwertigen Anlagen bestehen würden.

Die Wortbreite

Die Wortbreite oder auch Bit-Tiefe bestimmt wie eingangs erwähnt die Präzision, mit der analoge Eingangsspannungen am AD-Wandler nach der Digitalisierung und DA-Wandlung rekonstruiert werden können. Gleichzeitig bestimmt die Wortbreite jedoch auch, wie groß der Abstand zwischen der kleinsten möglichen Signalspannung (= dem leisesten Pegel) und der größten Signalspannung (= dem lautesten Pegel) ist. Damit bestimmt die Zahl der zur Verfügung stehenden Bits auch den Dynamikumfang eines digitalen Audio-Systems. Während ein 16-Bit-Signal eine theoretische Dynamik von 96 dB darstellen kann, erreicht ein 24-Bit-Signal eine Dynamik von 144 dB. Theoretisch deswegen, da man in der Praxis den kompletten Wertebereich nur selten voll ausnutzen kann. Denn sobald der maximal darstellbare Pegel mehr als drei Samples lang in Folge erreicht wird, wird die Wellenform „abgeschnitten“ und das Signal klingt nach der DA-Wandlung je nach Tonmaterial mehr oder weniger brutal verzerrt (Clipping). Es ist daher absolut vorrangig, keine solchen digitalen Clippings bei der Aufnahme zu erzeugen. Um dem vorzubeugen, lässt man in der Praxis großzügig Luft im Signalpegel (Head-

room). Die Folge davon ist jedoch, dass der durchschnittliche Signalpegel so niedrig wird, dass nur einzelne, sehr kurze Pegelspitzen (Peaks) den Wertebereich der zur Verfügung stehenden Bits annähernd nutzen. Die meiste übrige Zeit wird nur ein Bruchteil der möglichen Bit-Werte genutzt, wodurch z. B. ein 16-Bit-Signal netto oft nur noch 12-Bit-Qualität hat. Daher sollte man Einzelspuren oder Arbeitsdateien tunlichst mit 24 Bit Wortbreite aufzeichnen, sodass auch bei vorsichtiger Aussteuerung noch ein sehr guter Netto-Dynamikumfang bzw. eine sehr gute Präzision zur Verfügung steht und weniger Rundungsfehler bei der Effektbearbeitung zu erwarten sind. Zusätzlich sollte man überlegen, einzelne Pegelspitzen von einem analogen Kompressor oder Limiter im Aufnahme-

Unharmonische Verzerrungen nehmen in einem digitalen System mit sinkendem Pegel und steigender Frequenz zu.

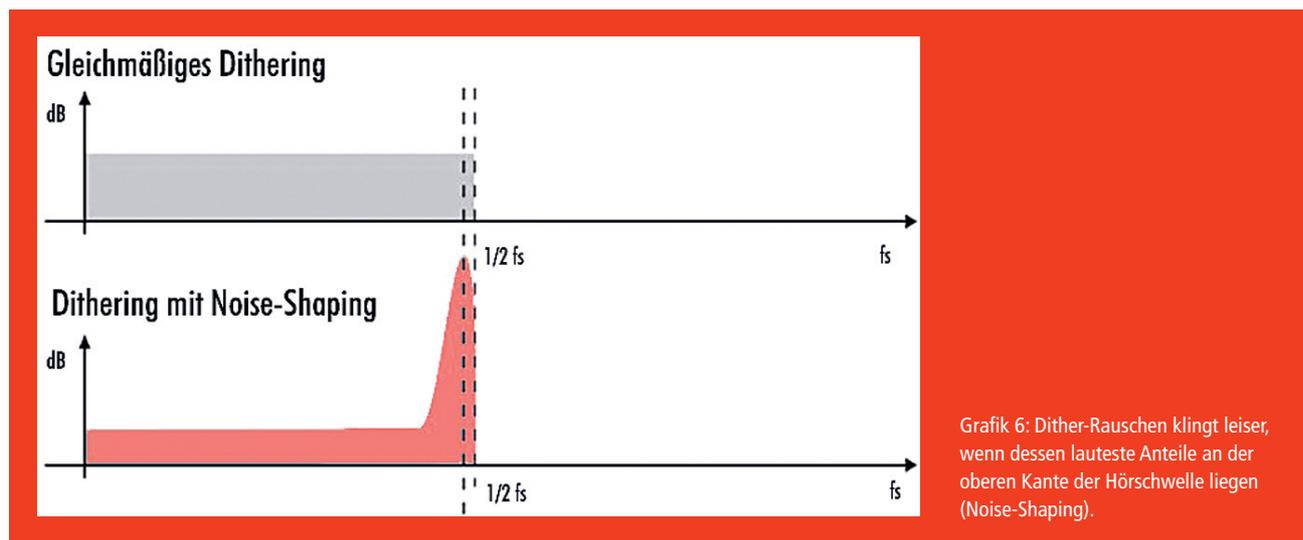
weg so weit absenken zu lassen, dass der Durchschnittspegel am AD-Wandler gefahrlos auf ein Niveau von ca. -16 bis -12 dB angehoben werden kann. Der rechnerische Signal-Rauschabstand überragt den Dynamikumfang eines 24-Bit-Signals um weitere 2 dB, was uns jedoch real wenig nützt, da die besten aktuellen Frontends (Mikrofon + Vorverstärker + Wandler) bestenfalls einen Signal-Rauschabstand von rund 115 dB erreichen. Im Gegensatz zu verschiedenen Samplingraten ist es in Recording-Programmen in der Regel kein Problem, 16- und 24-Bit-Dateien im selben Projekt parallel zu verwenden.

Intersample-Peaking

Egal ob 2 Bit, 8 Bit oder 24 Bit: Der positive oder negative Maximalwert eines Digitalsignals wird intern mit 0 dB FS (Full-Scale) referenziert; daher haben alle „guten“ Pegelwerte bei Digital-Audio ein Minus davor. Alles über 0 dB FS wird gnadenlos beschnitten und erklingt verzerrt. Die meisten digitalen Pegelanzeigen zeigen eine Übersteuerung an (Overload/Clip), wenn drei oder mehr Samples in Folge 0 dB FS entsprechen. Leider kann es passieren, dass auch Samplefolgen unter 0 dB FS am Ausgang des DA-Wandlers analoge Pegel von bis zu +8 dB erzeugen, welche die nachfolgenden Stufen in der Regel rettungslos überfordern. Dies liegt an der Art und Weise, wie gängige Wandler die Sample-Daten interpolieren, um analoge Wellenformen zu erzeugen. Um zu vermeiden, dass eine Audiodatei tatsächlich niemals Pegel oberhalb von 0 dB FS erzeugt, braucht man eine Pegelanzeige, welche ein mehrfaches Oversampling zur Analyse durchführt; erst dann werden auch diese so genannten Intersample-Peaks zuverlässig angezeigt. Eine solche Anzeige können Sie kostenlos unter www.dynamicrange.de herunterladen. Um beim Aussteuern einer Mischung nicht nach jeder Bearbeitung den kompletten Song mit der Pegelanzeige komplett in Echtzeit checken zu müssen, empfiehlt sich die Verwendung eines Oversampling-Limiters wie z. B. des Voxengo Elephant und die strikte Einhaltung eines maximalen Ausgangspegels von höchstens -0,5 dB.

Quantisierungs-Verzerrungen

Verzerrungen? Das macht doch nichts! Wozu all die Röhren-Emulationen, Bandsättigungs-Algorithmen und anderen mutwilligen Klangverschlechterer, die wir ständig freiwil-



Grafik 6: Dither-Rauschen klingt leiser, wenn dessen lauteste Anteile an der oberen Kante der Hörschwelle liegen (Noise-Shaping).

lig benutzen? Nun, es handelt sich bei diesen um mehr oder weniger gekonnte Nachbildungen analoger Systeme. Und diese verzerren auf ganz andere Weise als digitale Systeme, nämlich in Form von Oberwellen, welche in festen Intervallen zum Grundton der aktuellen Schwingungs-Folge stehen. Man nennt diese Abweichungen daher auch harmonische Verzerrungen. Der Grad solcher Verzerrungen steigt mit dem Eingangspegel an, was unserem biologischen Hörverhalten entspricht. Im psychoakustischen Umkehrschluss setzen wir **Verzerrungen** intuitiv mit hoher Lautstärke gleich, da unser Gehirn das Gehörte mit „mehr Schalldruck“ interpretiert. Digitale Systeme verhalten sich entgegengesetzt: Je kleiner der Pegel, desto kleiner wird die Dynamikauflösung und desto mehr daraus resultierende Verzerrungen sind mess- und hörbar. Man nennt sie Quantisierungs-Verzerrungen, welche klanglich meist mit den Begriffen „Rauheit“ und „Schärfe“ beschrieben werden. Diese Verzerrungen nehmen mit steigendem Signalpegel ab (!) und schlagen erst beim Erreichen des Maximalpegels ohne Vorwarnung in hässliche unharmonische Obertonverzerrungen um (Clipping). Ein analoges System ist da großzügiger: Lange bevor störende Verzerrungen einsetzen, machen sich subjektiv eher angenehme Sättigungs-Effekte bemerkbar und kündigen das Zuviel an Pegel zunehmend „drohender“ an.

Dithering

Mindestens so wichtig wie der Umgang mit den Pegelspitzen ist jedoch die Problematik rund um die allerleisesten Signalanteile. Einer der gravierendsten Unterschiede zwischen analogen und digitalen Audiosignalen ist nämlich deren Verhalten bei sehr kleinen Pegeln. Ein Analogsignal kann selbst dann noch deutlich wahrgenommen werden, wenn dessen Pegel geringer ist als der des Rauschteppichs (Noise-Floor). Dies kann man z. B. am Ende eines Fade-Outs sehr gut hören: das Signal „versinkt“ im Rauschen. Ein digitales Audiosignal hört dagegen abrupt auf, sobald der kleinste mögliche Wert erreicht ist. Bei einem rohen 16-Bit-Signal ist die Differenz zwischen dem kleinsten und dem zweitkleinsten binären Wert noch gerade so hörbar: das Signal scheint hart in den Rauschteppich „umschalten“. Eine Erhöhung der Bit-Tiefe auf z. B. 24 Bit minimiert die Wahrnehmbarkeit dieses Verhaltens zwar, das Grundproblem bleibt jedoch bestehen. Zudem steht spätestens vor dem Brennen auf CD die Reduktion auf

16 Bit ins Haus, sodass das Problem beim Abspielen derselben wieder voll zum Tragen käme. Eine naheliegende Möglichkeit, das unelegante Abbrechen des Signals zu maskieren, ist es, einige der niederwertigsten Signal-Bits (Least Significant Bits oder kurz LSB's) nach dem Zufallsprinzip an- und auszuschalten. Dadurch wird ein künstliches digitales Rauschen erzeugt (Dithering), in welchem das Signal sodann – wie bei analogen Aufnahmen – „versinken“ kann. Der offenkundige Nachteil dieser Strategie ist, dass sich der Signal-Rauschabstand verschlechtert – mutwillig.

Stille Wandler rauschen hoch

Doch auch diesem Missstand können man etwas entgegensetzen: Findige Ingenieure haben erkannt, dass Dither-Rauschen auch dann die leisen Signalanteile noch angenehm verpackt, wenn dieses vorrangig zwischen der oberen Hörschwelle bis hin zur Nyquist-Frequenz von $\frac{1}{2} f_s$ liegt (Noise-Shaping). Oberhalb von ca. 18 kHz ist unser Gehör nicht mehr sehr empfindsam gegenüber Rauschen, sodass die subjektive Lautheit des Dither-Rauschens deutlich niedriger ausfällt (siehe Grafik 6). Um den Balanceakt zwischen der Maskierungsfähigkeit des Ditherings bei kleinen Pegeln und der Wahrnehmbarkeit des hinzugefügten Rauschens zu meistern, versucht jeder Hersteller auf seine Art den Rausch-Frequenzgang im Höhenbereich zu optimieren. Die beiden populärsten Dithering-Algorithmen bzw. besser gesagt Algorithmus-Pakete sind UV22(-HR) der Firma Apogee und POW-R des POW-R-Konsortiums, welches aufgrund der verschärften Lizenzierungs-Bedingungen von Apogee von verschiedenen Experten vor rund 10 Jahren ins Leben gerufen wurde. Beide Strategien zielen darauf ab, bei der Reduzierung der Bit-Tiefe von 24 auf 16 Bit so viel wie möglich von der Detailtreue der höheren Wortbreite hinüberzuretten und bei niedrigen Pegeln ein elegant geglättetes Abfallen in den Rauschteppich hörbar zu machen. Das Funktionsprinzip von UV22 ist, dass vorab gespeicherte Dither-Muster pegel- und frequenzabhängig aus einer festen Tabelle abgerufen und in das Signal eingespeist werden. POW-R kommt in drei Varianten, welche jeweils für Sprache, Klassik und Pop/Rock optimiert wurden. Hier wird schmalbandiges Dither-Rauschen unterhalb der Nyquist-Grenze eingefügt, doch die Entwickler legen Wert darauf zu betonen, dass es sich nicht einfach um Noise-Shaping handelt. Experten wie Bob Katz konnten auf

16 Bit geditherte POW-R-Dateien gehörmäßig nicht mehr vom 24-Bit-Original unterscheiden; der Algorithmus scheint dem betagten UV22 den Rang abgelaufen zu haben. Doch für solche Feinheiten muss man das Gras wachsen hören, denn selbst unter gehobenen Abhörbedingungen wird das Dithering beider Algorithmen psychoakustisch so unauffällig und „musikalisch“ eingefügt, dass der Rauschanteil nicht hörbar ansteigt und die Differenziertheit des 24-Bit-Originals subjektiv voll erhalten bleibt. Die Folgen solch genialen Ditherings für unser Verständnis von digitalen Audiosystemen reichen so weit, dass wir den Begriff der „Auflösung“ nicht mehr ausschließlich über die Wortbreite definieren sollten: In der Praxis bestimmt heute die Qualität des Ditherings die Auflösung eines Digital-Audio-Signals, während die Bit-Tiefe lediglich dessen Dynamikumfang definiert.

Erkenntnisse

Wer bis hierhin alles aufmerksam gelesen hat, wird vielleicht besser verstehen, dass „hohe Zahlen“ kein Garant für eine hohe Klangqualität sind: Eine gut geditherte 16-Bit-Datei bei 44,1 kHz Samplingrate kann durchaus wesentlich besser klingen, als eine zu niedrig angesteuerte, verjitterte 24-Bit-Datei mit 192 kHz Samplingrate. Die Gute Nachricht ist, dass wir trotz der vielen systembedingten Klangverschlechterungen, die Digital-Audio mit sich bringt, heute schon mit sehr preiswerten Mitteln einen Sound produzieren können, von dem man vor wenigen Jahren nur träumen konnte. Allerdings nur dann, wenn man die Prioritäten bei der Aufnahme richtig setzt. Um dies optimal zu bewerkstelligen, empfehlen wir Ihnen den Beitrag „Die Soundpyramide“ in diesem Heft. ■

▶▶ Weiterführende Links

1. Intersample-Peaking u.a.: www.dynamicrange.de/de/de/faqtech-info
2. .POW-R: www.mil-media.com/pow-r.html
3. Apogee UV22: www.apogeedigital.com/pdf/UV22-OS.pdf
4. Bob Katz zum Thema Dithering: www.proaudioreview.com/article/924

Anmerkung: Leider sind die Quellen 2.-4. nur auf Englisch verfügbar.