SAMPLING IN DIGITALEN AUDIOSYSTEMEN

von Roman Groß - Diplom-Informatiker FH

Roman Groß - New Perspectives On Sound
Frühjahr 2012

INHALTSVERZEICHNIS

Historie

Dr. Harry Nyquist (Aussprache: [nʏːkvɪst], nicht [naɪkwɪst]), US-amerikanischer Physiker an der Yale University entdeckte 1927, als er bei den Bell Laboratories arbeitete, seine Samplingtheorie. Seine Forschungsergebnisse publizierte er 1928 unter dem Titel *Certain Topics in Telegraph Transmission Theory.* heute besser bekannt als *Nyquist-Shannon-Abtasttheorem.*

Nyquist Abtasttheorem: Eine gesampelte Signalform enthält die vollständige Information ohne jede Verzerrung, wenn die Abtastrate mehr als 2x so groß ist wie die höchste Frequenz der gesampelten Signalform.

Einführung

Dieser Artikel erklärt die Methode des Sampling im Allgemeinen. Sinn des Artikels ist es, weit verbreiterte Missverständnisse hinsichtlich einer Audio Sampling-Rate von 192 kHz zu beseitigen. Diese Missverständnisse beruhen auf falschen Voraussetzungen, die von Marketing-Strategen erdacht wurden.

Dass *Mehr* besser ist, entspricht dem gesunden Menschenverstand. So wie mehr Pixel ein höher aufgelöstes Bild ergeben oder schnellere Taktung einen schnelleren Computer soll der Eindruck erweckt werden, dass schnelleres Sampling zu höherer Auflösung und mehr Details führe.

Diese Analogie ist falsch. Die große Erkenntnis in Nyquists Theorem liegt darin, dass die vollständige Information vorliegt, mit 100% Auflösung und ohne Verzerrung, ohne die Notwendigkeit schnelleren Samplings.

Nyquist zeigte, dass die Samplingrate nur größer sein muss als die doppelte Signalbandbreite. Wie groß ist nun die Signalbandbreite bei Audiosignalen? Die Forschung hat gezeigt, dass Musikinstrumente durchaus Energie oberhalb 20 kHz liefern können, aber kaum oberhalb 40 kHz. Die meisten Mikrofone nehmen oberhalb 20 kHz nichts auf und das menschliche Gehör hört oberhalb 20 kHz nichts. Daraus folgt unmittelbar, dass die Audio-Samplingfrequenz der CD, 44,1 kHz ausreicht, die gewünschte Bandbreite von 0 Hz -20 kHz fehlerfrei und unverzerrt darzustellen.

Sampling mit 192 kHz erzeugt größere Dateien, die mehr Platz benötigen, die Übertragungsgeschwindigkeit wird dadurch reduziert. 192 kHz erzeugt deutlich erhöhte Erfordernisse an die Rechenleistung und Rechengeschwindigkeit der Prozessoren. Zwischen Rechengeschwindigkeit und Rechengenauigkeit besteht ein Zusammenhang. Geschwindigkeitsbedingte Ungenauigkeiten sind bedingt durch die Hardware Realitäten wie das Laden/Entladen von Kondensatoren, Einschwingvorgänge und mehr. Diese Prozesse können nicht beliebig beschleunigt werden. Verlangsamung verbessert die Genauigkeit.

Wenn also bereits 88.2 kHz oder 96 kHz Sampling bedeutend schneller ist als für eine fehlerfreie Signaldarstellung der Audiobandbreite notwendig, warum dann 192 kHz Samplingrate? Hier und da wird das Argument eines besseren Impulsverhaltens angeführt, implizierend, dass besseres Impulsverhalten für eine bessere Raumdarstellung sowie einen „analogeren“ Klang sorgen soll.

Solche Behauptungen zeigen, dass die Grundlagen der Sampling Theorie nicht verstanden wurden. Ein Mikrosekundenimpuls mag ein Argument für ein Megehertz-Audiosystem sein .. das es nicht gibt. Nicht ohne gute Gründe sind Musikinstrumente, Mikrofone und Lautsprecher auf den menschlichen Hörbereich, nicht auf Megahertzbandbreite abgestimmt.

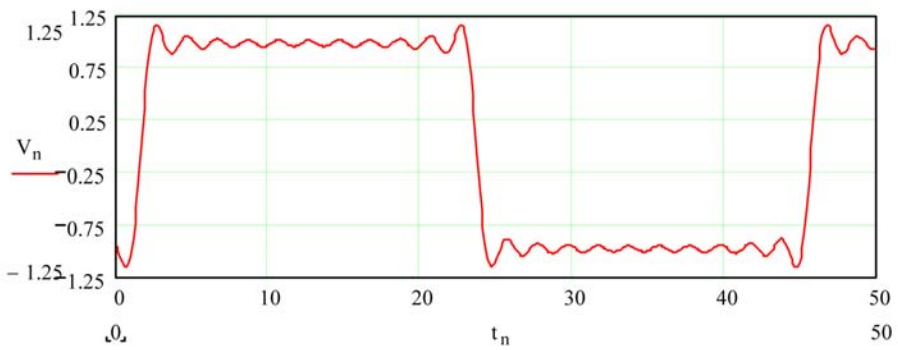
Die Audio Sample-Rate ist die Menge der Audio-Daten. Diese Daten werden generiert von einem A/D-Konverter (ADC) , wiedergegeben von einem D/A-Konverter (DAC). Das im Deutschen gebräuchliche Wort „Wandler“ ist eigentlich nicht korrekt, im Folgenden werden daher die Begriffe ADC und DAC verwendet.

Verwirrung mag stiften, dass einige lokale Prozesse der digitalen Audiodatenverarbeitung schneller ablaufen als die Sampling-Rate, So arbeiten die meisten Frontends moderner ADCs (Die Modulator-Sektion) 64-512x schneller als die Sampligrate von 44,1 kHz oder 48 kHz. Dies ist 16-128-fach schneller als 192 kHz. Diese Hochgeschwindigkeitsverarbeitung erzeugt nur ein paar Bits. Sie so generierten Bits werden einem sogenannten Dezimationsprozess zugeführt, der die Geschwindigkeits für mehr Bits reduziert. Dem lokalen Konverter (wenige Bits bei Mhz-Geschwindigkeit) folgt ein Dezimierungs-Schaltkreis, das die benötigten Bits der finalen Samplingrate erzeugt.

Beides, die Über-Alles-Sampling-Rate sowie die Prozessrate in bestimmten lokalen Schaltungen des ADC werden öfter als „Sample-Rate“ bezeichnet. Es ist zu unterscheiden zwischen der Audio Sample Rate und anderen Sample Rates, wie die Eingangsstufe eines ADC oder das Oversampling eines DAC.

Sampling

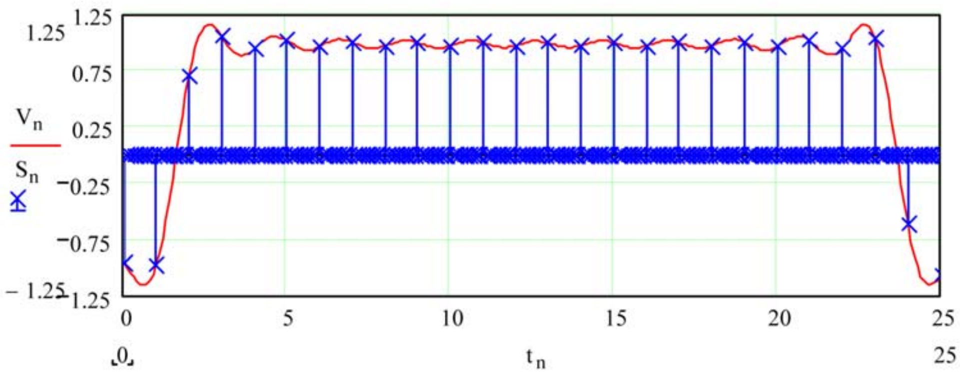
Sehen wir uns ein bandlimietiertes Rechtecksingnal an. Die Grundfrequenz beträgt 1 kHz und die Bandbreite 22,05 kHz, wie im Red Book der Audio CD. Die Berechnung ergibt 22 Harmonische, die Grafik zeigt die Addition dieser 22 Harmonischen:



Hier ein vergrößerter Ausschnitt der Kurve (rot) zwischen t=0 und t=25. Die blauen Linien zeigen die Sample-Zeitpunkte. Die blauen X zeigen die Werte, die den einzelnen Samples zugewiesen wurden.

Zum Beispiel ist der Wert des Samples bei t=1 ca. -1, bei t=2 haben wir 0,75 usw.

Die gesampelte Welle kann nun als Reihe von Werten dargestellt werden



Die spontane Reaktion ist gewöhnlich, dass wir hier nicht genügend Werte haben um die korrekte Wellenform (rot) in allen Details zu rekonstruieren. Die spontane Reaktion ist falsch.

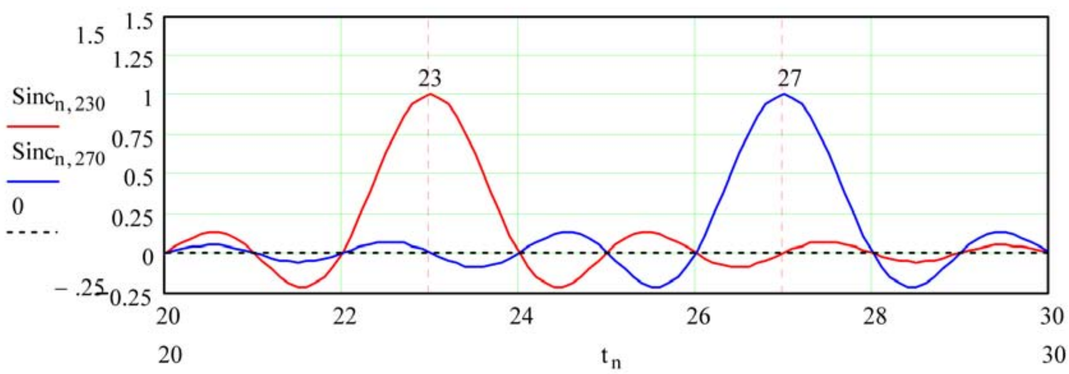
Der Schlüssel zum Verständnis liegt darin, dass das Signal bandbreitenlimitiert ist. Für eine gegebene Bandbreite muss die Anzahl der Samples nur 2x so groß sein wie die Bandbreite um die korrekte Wellenform **originalgetreu** zu rekonstruieren, inklusive aller Feinheiten zwischen den X´en.

Schauen wir uns an, wie das geht …

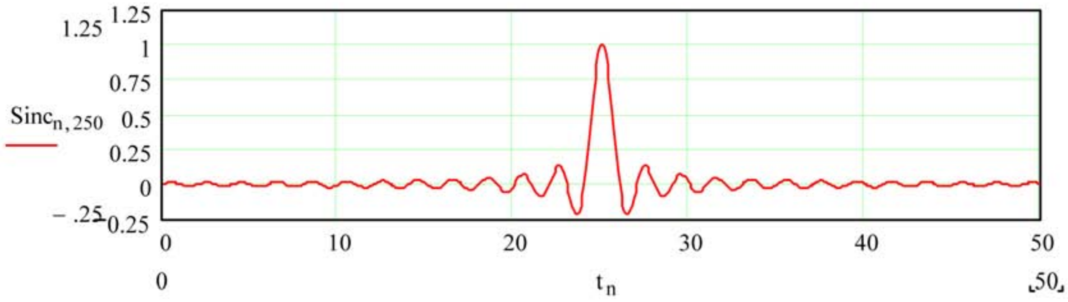
Dazu muss der Sinus Cardinalis *sinc* vorgestellt werden. Für die mathematisch Interessierten: sinc(x)=sin(x)/x. Zeichnen wir zwei sinc Funktionen, eine zentriert auf die 23 (rot) und die zweite bei 27 (blau).

Jede sinc-Kurve enthält die Hauptwelle und sog. „Ringing" (abklingende Sinuskurve) beidseits der Hauptwelle. Beachte: Alle sinc´s haben dieselben Nulldurchgänge (schwarze gestrichelte Linie), bei t=20,21,22,23 …

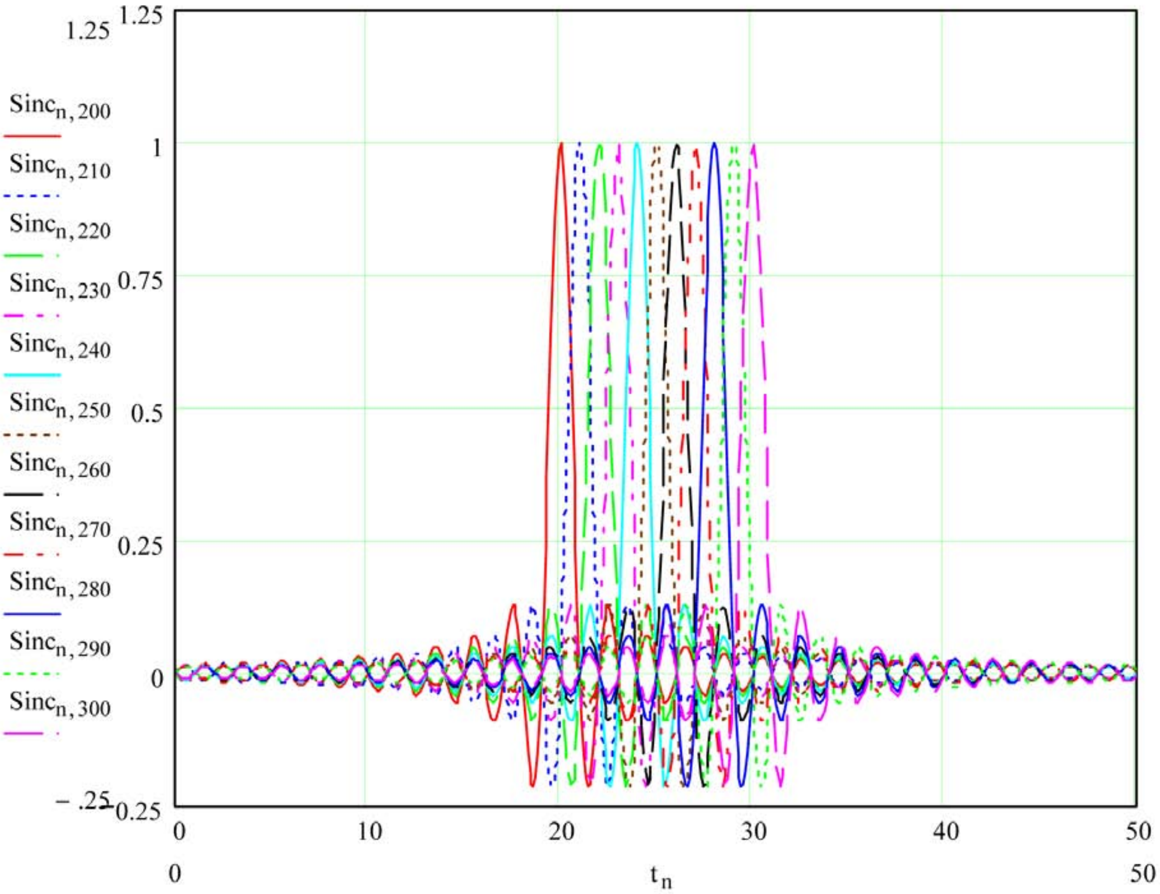
Alle unsere sinc-Funktionen werden so positioniert, dass sie an den selben Stellen nullen (Nulldurchgang). Beachte, dass das Ringing zwischen den Peaks der beiden sincs (zw. 23 und 27) in entgegengesetzte Richtungen geht. Das Ringing unterhalb 22 und oberhalb 28 geht in dieselbe Richtung.



Wir sehen, dass der sinc plot mit dem Ringig fortfährt über die gewählten Grenzen des Plots (20 bis 30) hinaus. Erweitern wir den angezeigten Bereich, sehen wir, dass das Ringing weitergeht, aber in der Amplitude kleiner wird, je weiter wir uns vom Zentrum entfernen. Je weiter wir uns entfernen, tendiert es gegen Null,

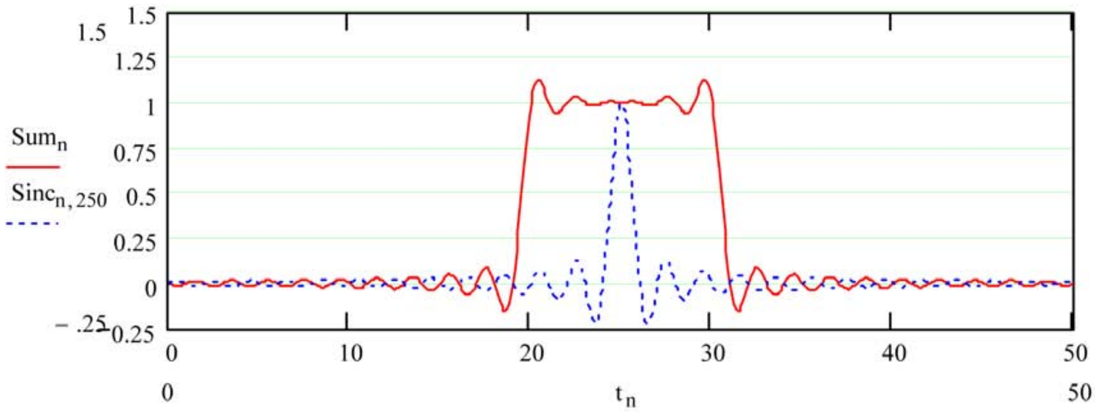


Nun generieren wir 11 sinc Funktionen gleicher Amplitude und verteilen sie. Wir zentrieren sie bei T=20,21,22,23 .. usw.

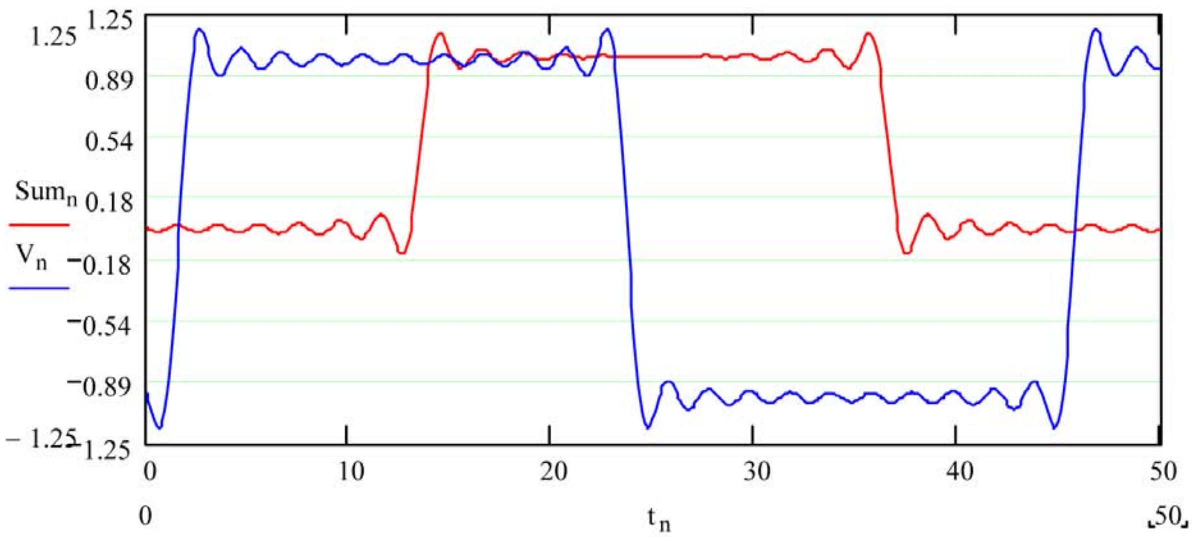


Die Addition der 11 amplituden-identischen sinc-Funktionen ergibt die hier dargestellte Wellenform. Wir erkennen hier eine gewisse Ähnlichkeit zum bandlimitierten Rechtecksignal.

Die blaue Kurve zeigt zum Vergleich einen einzelnen sinc.

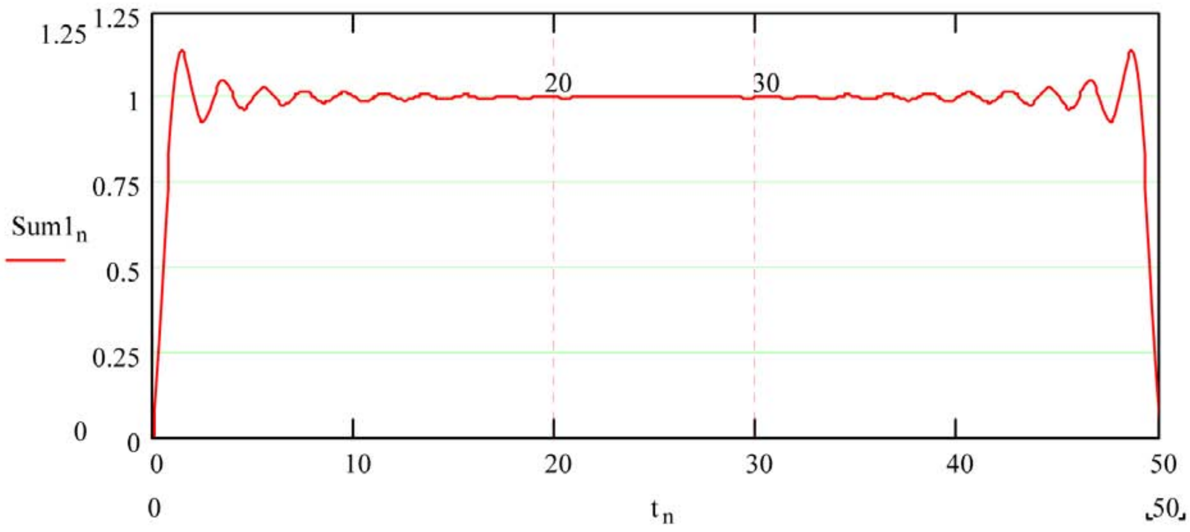


Nun addieren wir 22 sinc´s nach demselben Verfahren. 22, weil dies die Hälfte eines 1 kHz Rechtecksignals ergibt. Die rote Kurve zeigt das Ergebnis, die positive Hälfte eines bandlimitierten Rechtecksignals. Die blaue Kurve zeigt das ursprüngliche Signal, siehe Fig.1. Ignoriere den Zeitversatz in der Darstellung.

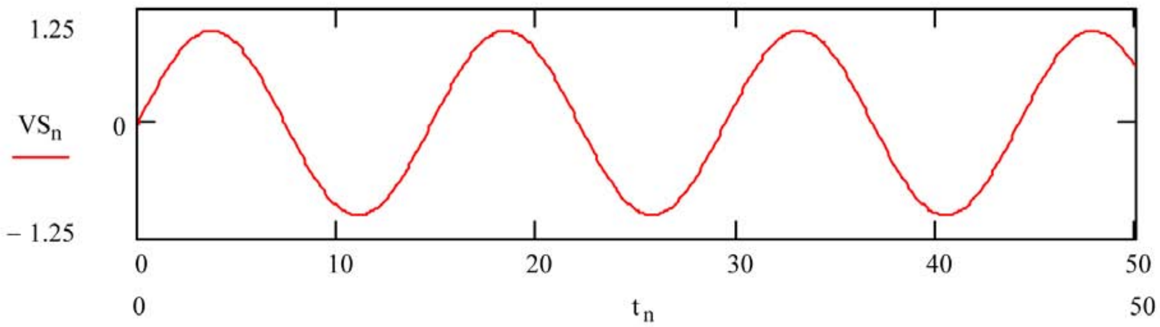


Kann man durch sinc-Funktionen auch Gleichspannung (DC) generieren? Der nächste Plot zeigt die Summe von 480 sinc´s gleicher Amplitude. Wie vorher dargestellt, sind die Positionen der sinc´s so gewählt, so dass die Spitze der Hauptwelle und die Nulldurchgänge des Ringings zum Samplezeitpunkt auftreten. Wie man sieht, ist der mittlere Bereich der Summe (t=20 bis t=30) eine gute Annäherung eines DC-Signals. Jedes Einzelsignal ist weit weg von DC, aber die Summer ergibt DC.

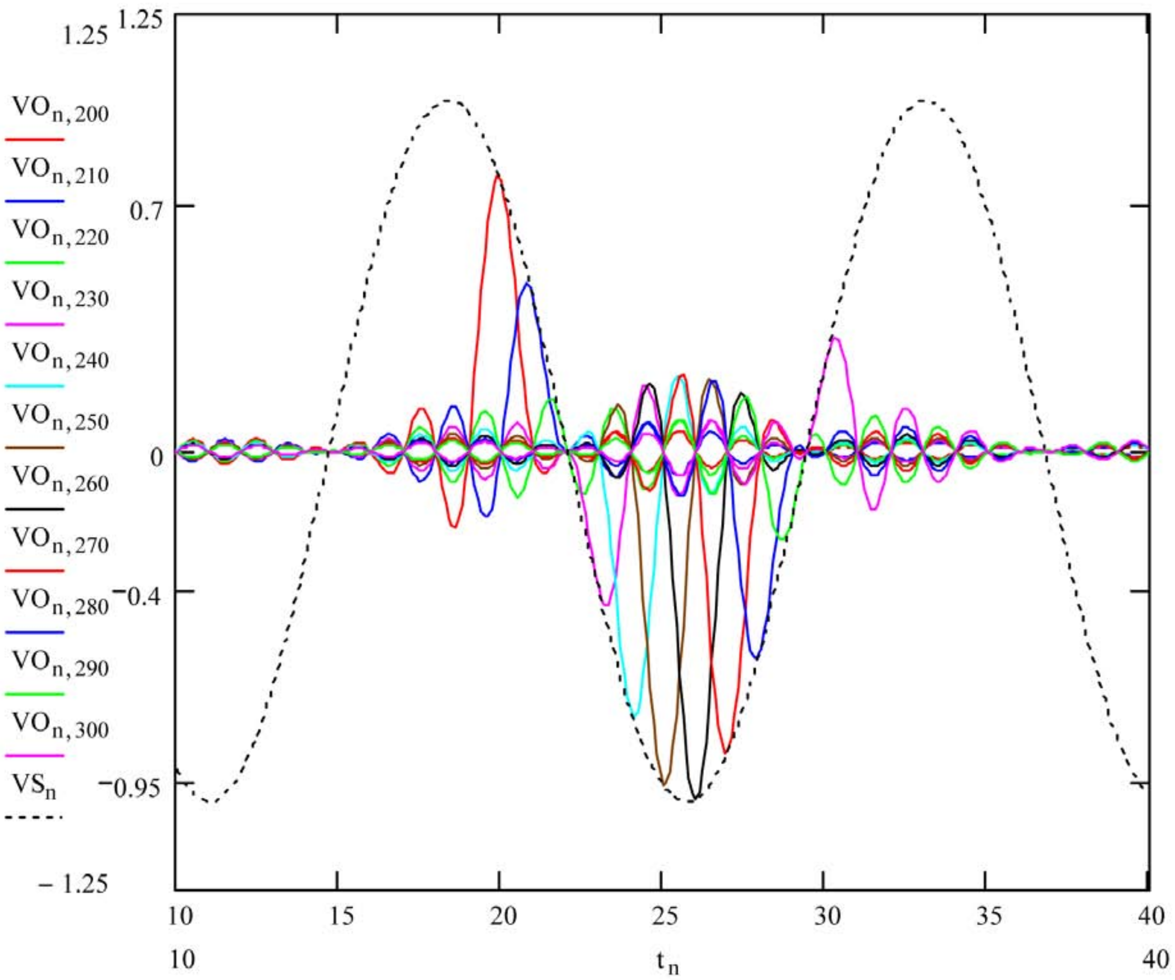
Zu jedem beliebigen bandbreitenlimitierten Signal (also jedem möglichen Audiosignal) gibt es einen Satz sinc-Funktionen, deren Summe dem gewünschten Signal entspricht-



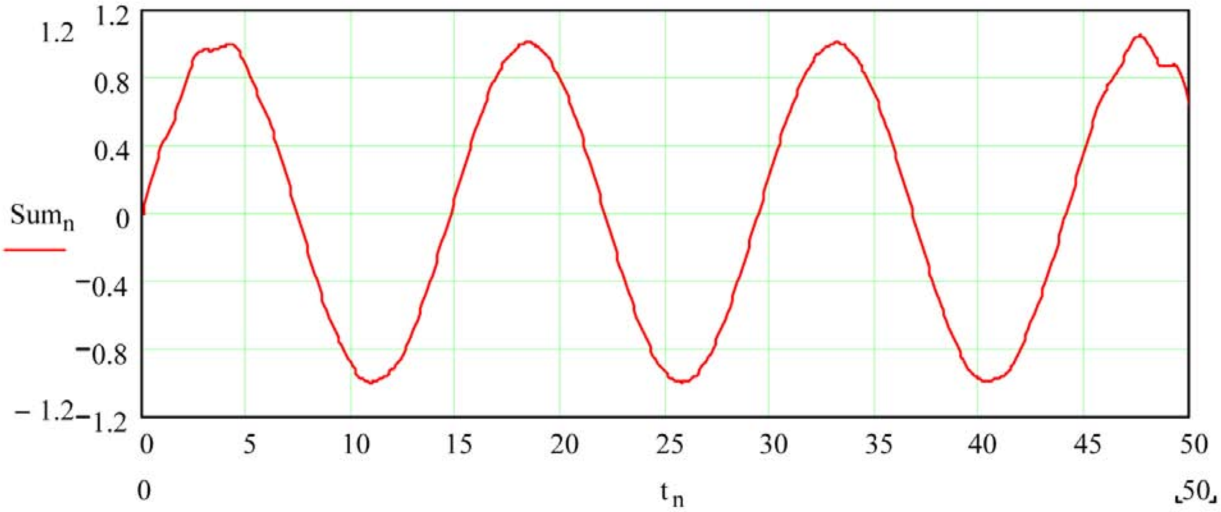
Suchen wir den sinc-Funktionen-Satz zu einem 3 kHz-Sinuston. Hier der Plot des Signals:



Nun multiplizieren wir jeden Sample-Wert mit einer sinc-Funktion. Der Plot zeigt zeigt die Multiplikation des 3 kHz-Tons (schwarz gepunktet) mit einem sinc an 11 Stellen (20 bis 30). Das Zentrum jedes sinc ist synchronisiert und eingepegelt auf den jeweiligen Sample-Wert. Für negative Werte des Sinus ist die sinc-Funktion invertiert.



Mit den so angeordneten sinc´s und multipliziert mit dem Einganssignal aller Sample-Punkte, sehen wir hier das Ergebnis. Alle sinc´s addiert ergeben dieses Bild:



Man sieht, dass die erste Halbwelle und die letzte verzerrt sind. Der mittlere Bereich (t=7 bis 45) sieht dem 3 kHz Eingangssignal ähnlich. Die Verzerrungen liegen am aprupten Start und Stop der Welle. Jeder plötzliche Start oder Ende führt zu Hochfrequenz-Inhalten.

Ein Sinus-Burst enthält die eigentliche Sinuswelle und eine Gating-Welle (die den Sinus triggert). Das Gate erfordert unendliche Bandbreite.

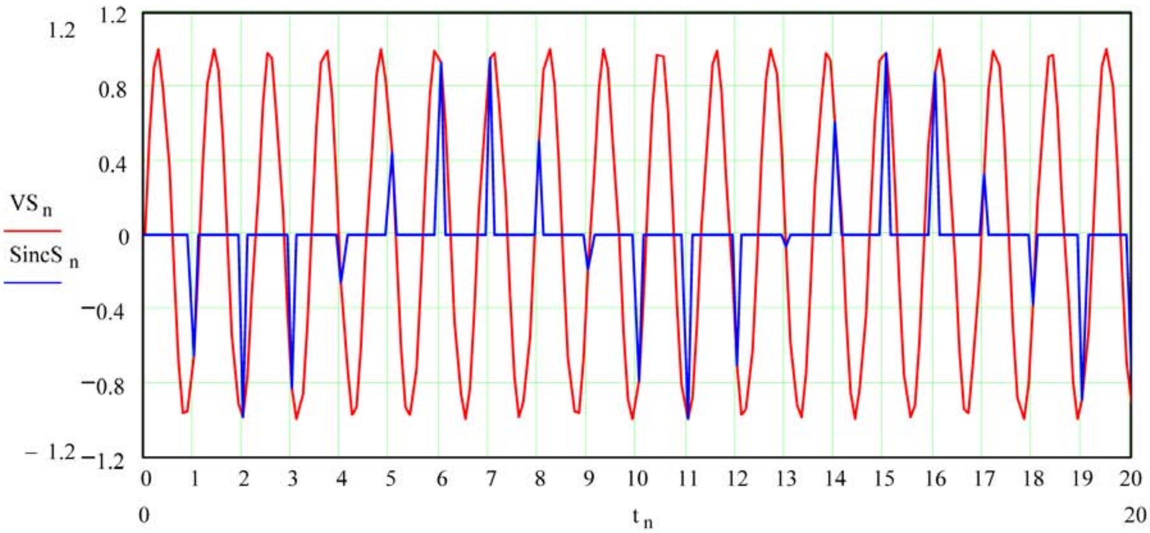
Jeder Samplepunkt ist das Ergebnis aller sincs, hier im Beispiel sind es 50 sinc´s. Das Red Book CD Format basiert auf 44100 sinc´s pro Sekunde!

Bedeutet das, dass die ersten paar Samples auf einer CD verzerrt sind? Nein, denn der Grund für die Verzerrung ist hochfrequenter Inhalt des Bursts. Ohne hochfrequente Inhalte verschwindet die Verzerrung komplett.

Was ist hochfrequenter Inhalt? Hochfrequenz ist jede Frequenz oberhlab der Ringingfrequenz der sinc Funktion. Die sinc Ringingfrequenz muss höher sein als die zweifache Frequenz, die wir samplen wollen. Das Brechen dieser Regel erzeugt Aliasing Verzerrungen.

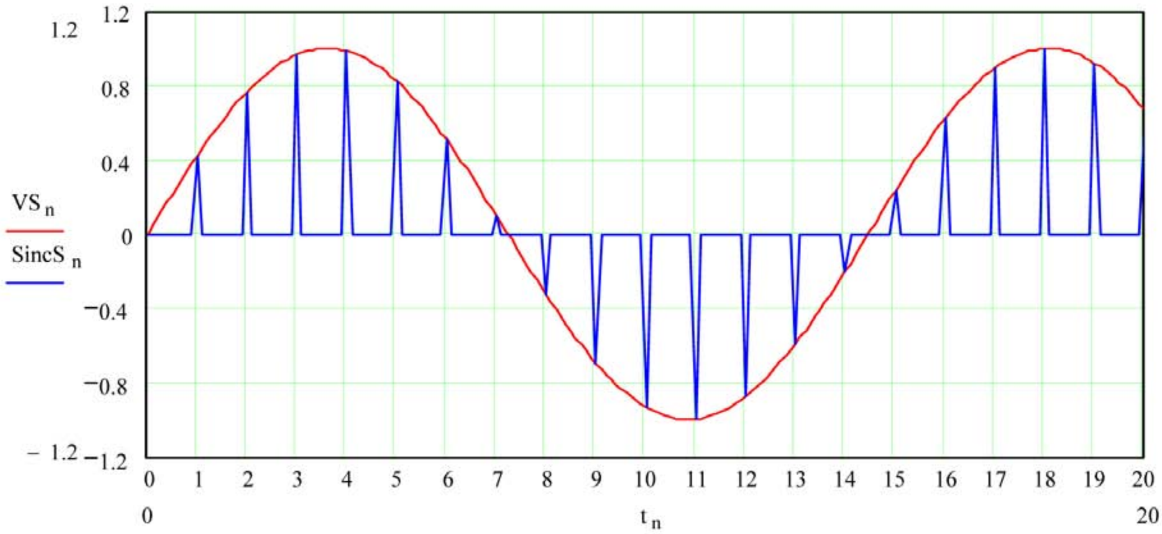
Dies wird durch die Bandbreitenbegrenzung des Signals auf 20 kHz zuverlässig verhindert, Schauen wir uns an, was passiert, wenn wir eine Frequenz oberhalb der Hälfte der Samplingfrequenz samplen. Die rote Welle ist die hochfrequente Eingangsfrequenz.

Blau zeigt die Samplepunkte und die Höhe der Sampling sincs. Wir haben nicht genug Samplepunkte um die Frequenz korrekt darzustellen. Die Summe der sincs wird ein falsches Resultat liefern. Daher ist mit einer Samplingrate von mehr als der doppelten Signalbandbreite gesampelt werden.



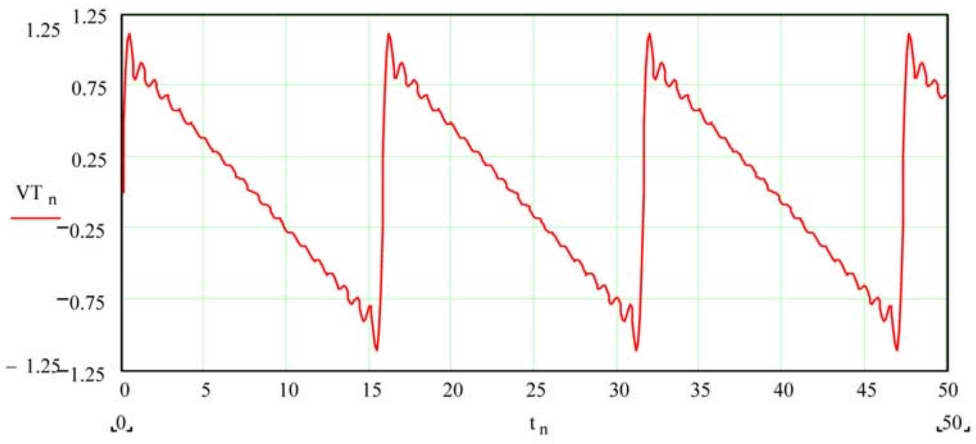
Im obigen Plot ist die Samplingrate zu langsam um die korrekte Wellenform darzustellen.

Hier eine Eingangswelle mit weniger als der halben Samplingfrequenz. Jetzt ist alles in Ordnung.

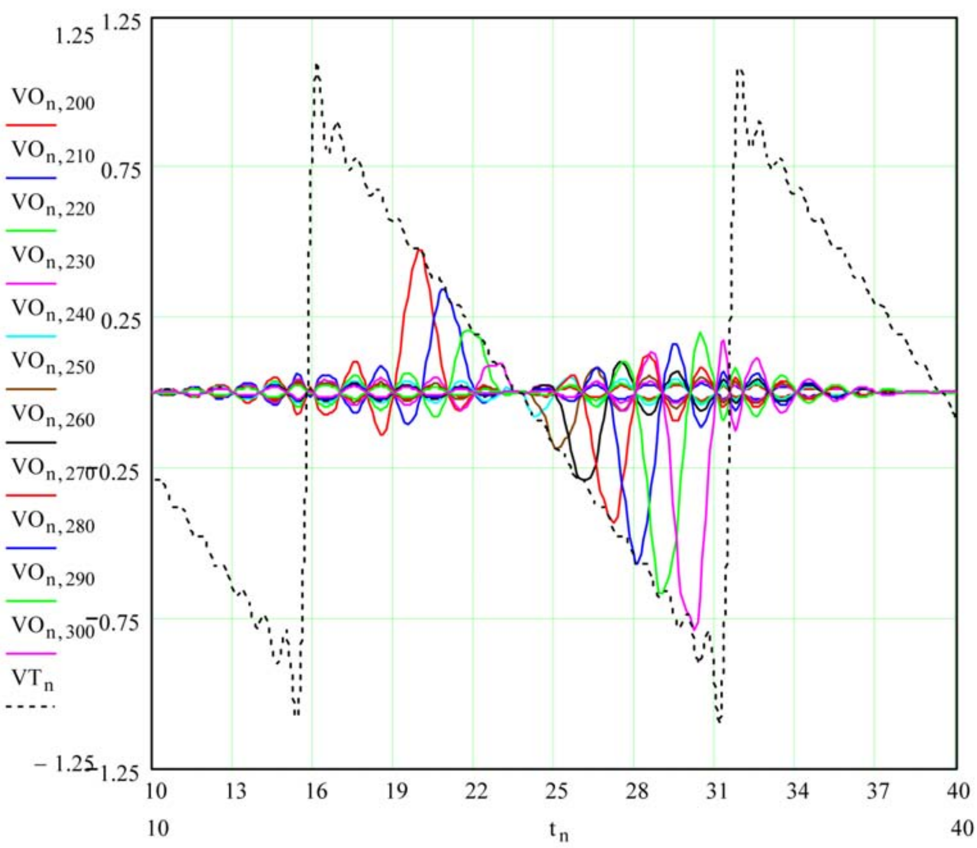


Untersuchen wir nun eine komplexere Signalform, ein bandbreitenlimitiertes Sägezahnsignal. Die unten dargestellte Welle wurde durch die Addition von 16 Harmonischen erzeugt. Alle Harmonischen oberhalb 22 kHz wurden auf Null gesetzt.

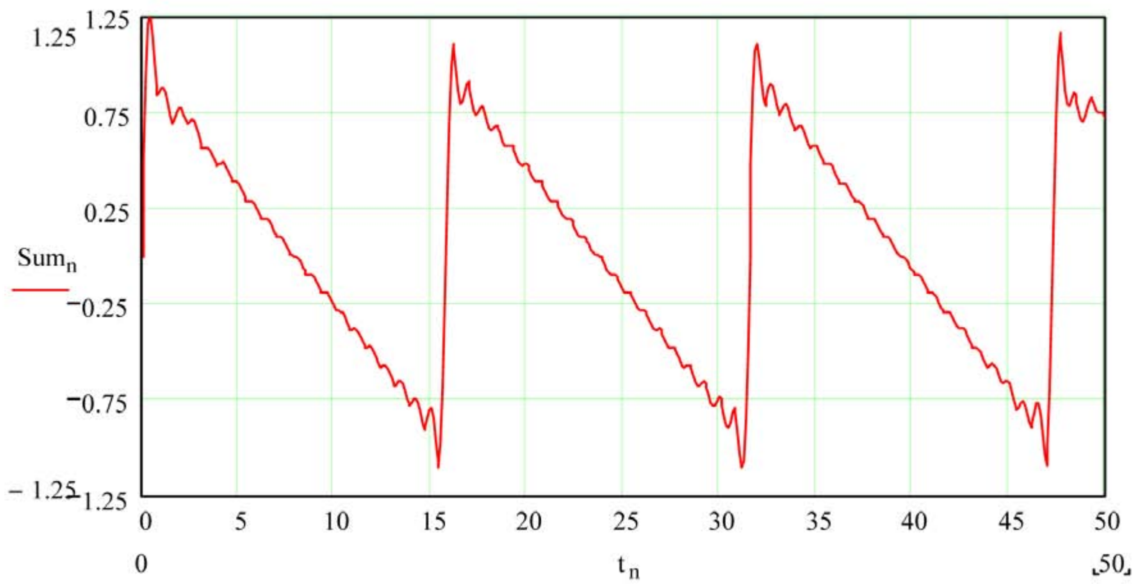
Ringing und die begrenzte Anstiegszeit des Signals sind die Folge der Bandbreitenlimitierung.



Hier die Darstellung der zugehörigen sincs



Die Addition aller sinc´s zeigt folgendes Bild. Am Anfang und Ende der Welle sehen wir wieder die weiter oben erläuterten Burst-Verzerrungen



Resümmée

Wir haben gezeigt wie jede bandbreitenlimitierte Wellenform aus sinc-Funktionen konstruiert werden kann. Wir haben gezeigt, dass es keinen Grund gibt, mit einer höheren Samplingfrequenz als der doppelten Bandbreite zu arbeiten, weil die Addition der sincs die korrekte Wellenform liefert.
Die Bandbreitenlimitierung ist die Garantie für perfekte Resultate, so wie nur 2 Punkte genügen, um eine Gerade zu definieren.

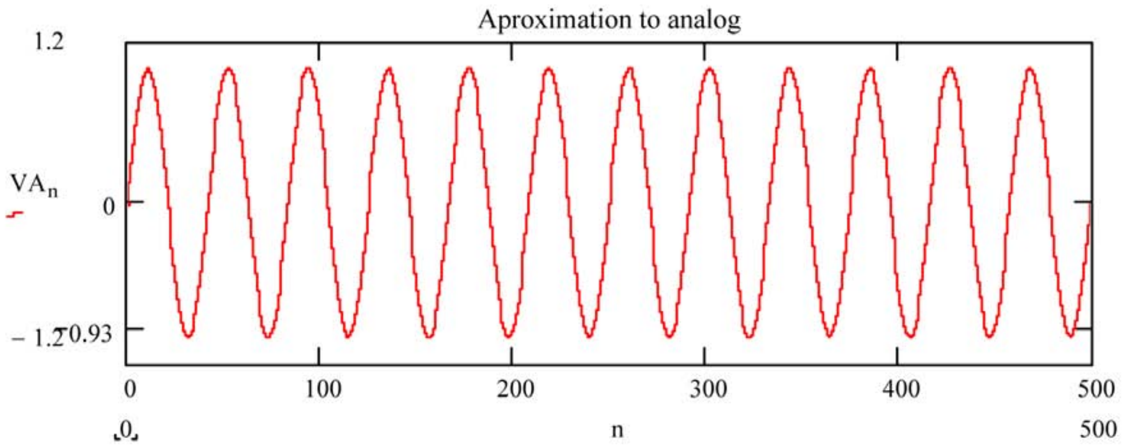
Wir benötigen also eine sinc-Frequenz, die das 2-fache der höchsten zu samplenden Frequenz übersteigt.

Filterung

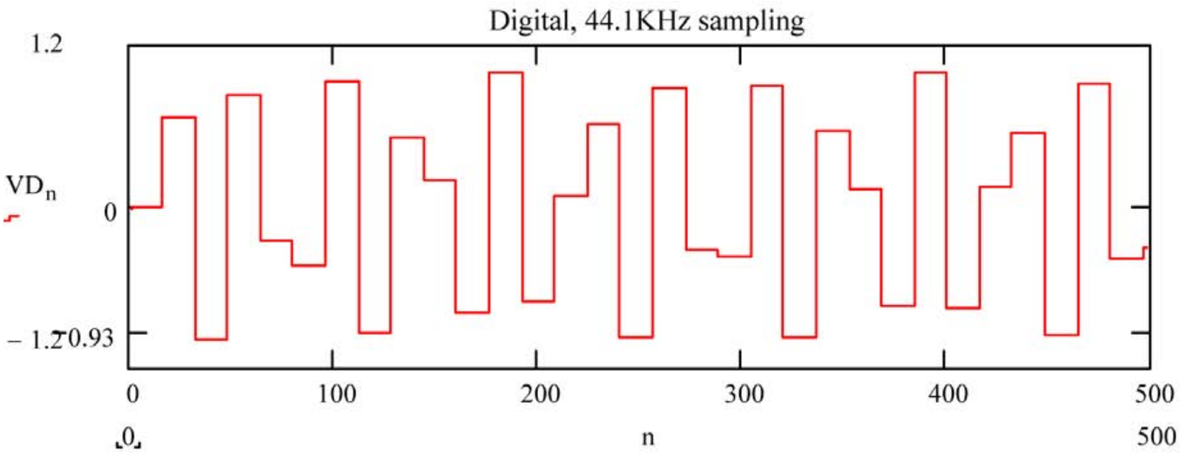
Für eine Bandbreite von 22,05 kHz werden nicht mehr als 44100 Samples benötigt. Eine Erhöhung der Samplerate erhöht die Genauigkeit der Signalrekonstruktion nicht.

Schauen wir uns dies grafisch an.

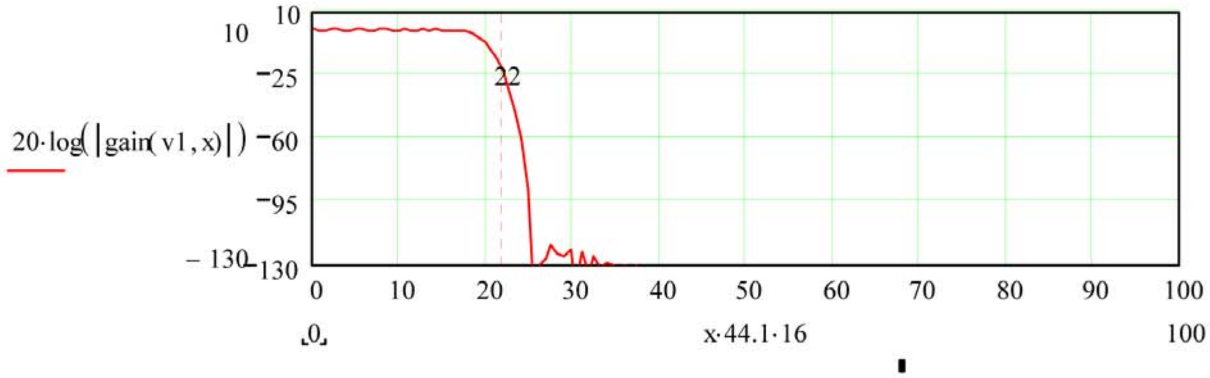
Wieviele Samplepunkte werden benötigt?
Ein Analogsignal ist kontinuierlich, weil es unendliche Werte enthält. Daher ist analoge Simulation nicht „computerfreundlich“. Zur Demonstration nehmen wir eine Anhäherung eines Analogsignals, gesampled mit 705,6 kHz, das 16-fache von 44,1 kHz.
Die Signalfrequenz ist 17 kHz.
Das Ergebnis sieht dem Originalton recht ähnlich.



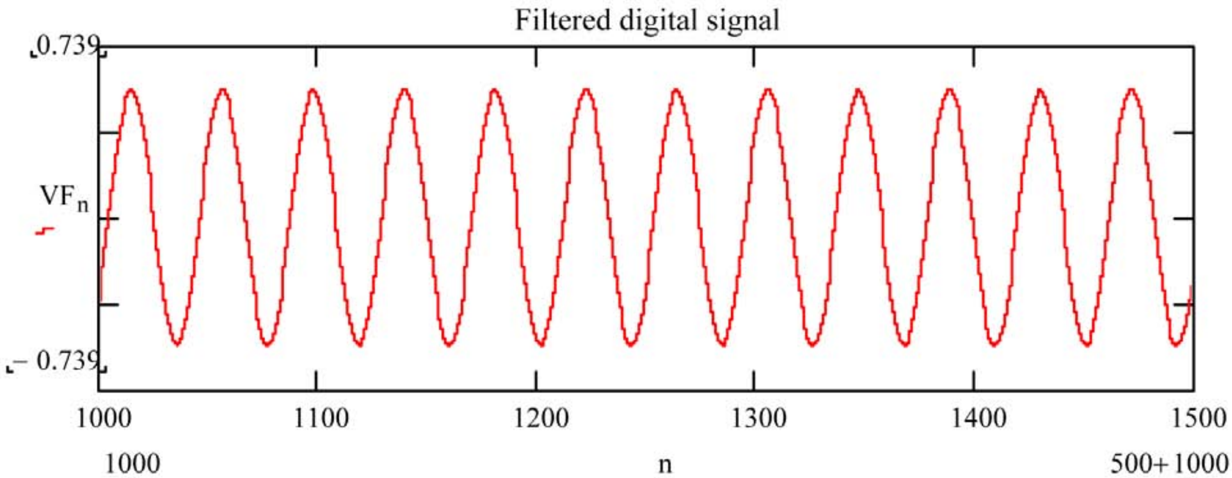
Hier ein mit 44,1 kHz gesampleter 17 kHz-Ton, man sieht, dass das Ergebnis einer Sinuswelle nicht sehr ähnlich sieht.



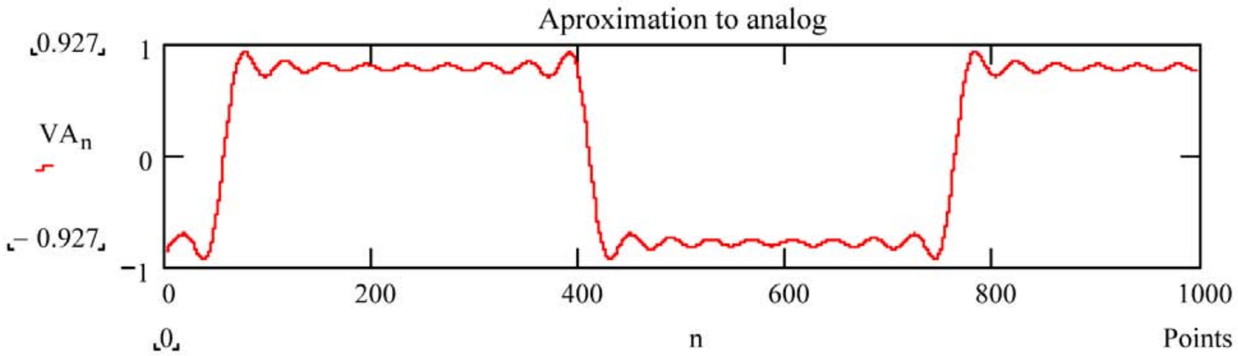
Filtern wir das obige Digitalsignal. Hier die Grafik der Filtercharakteristik:

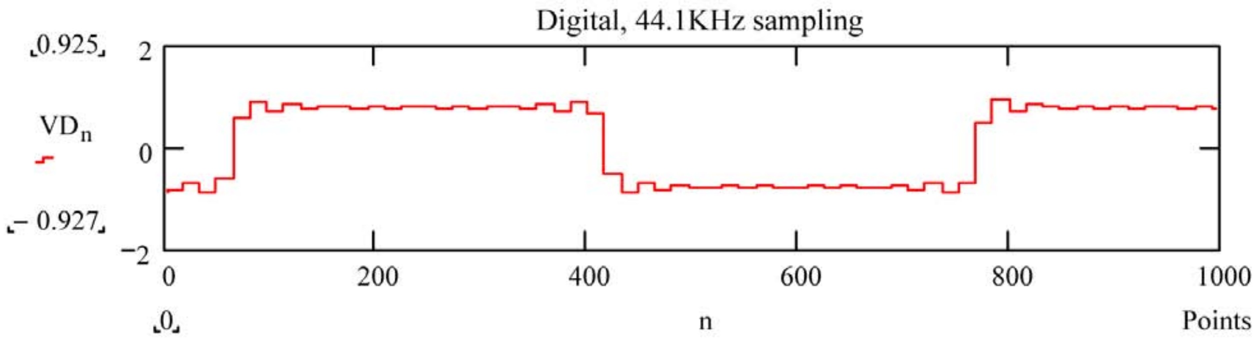


Hier das Ergebnis nach Filterung. Wie man sieht, ist der 17 kHz Sinus perfekt rekonstruiert.

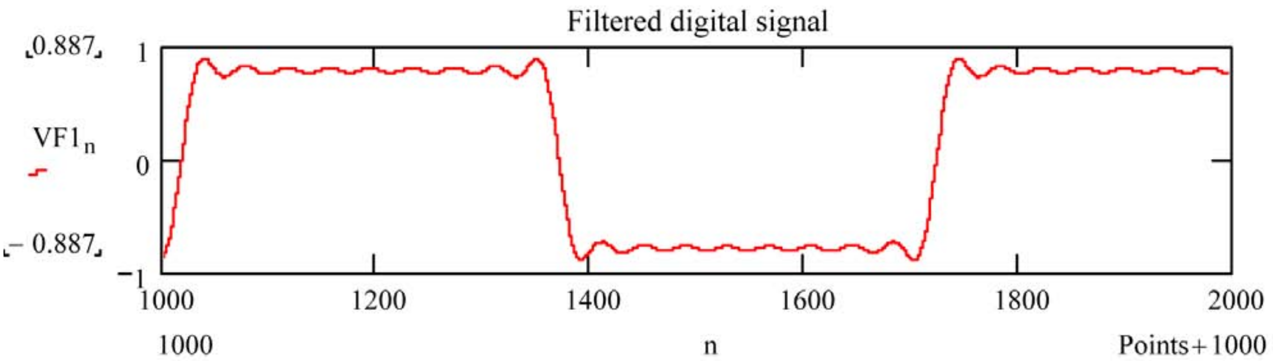


Wiederholen wir das Ganze mit einer 1 kHz Rechteckwelle. Das Rechtecksignal ist nicht „perfekt“, weil es auf 20 kHz bandbreitenlimitiert ist.





Wieder filtern wir das digitale Signal. Das Ergebnis sieht so aus:



Die wichtige Erkenntnis ist, dass im Endresultat ALLE Werte korrekt sind, nicht nur zum Sample-Zeitpunkt, sondern zu JEDER Zeit.

MEHR Sample-Punkte werden nicht benötigt. In höheren Samplingraten ist NICHT MEHR INFORMATION enthalten.

Sampling nach Nyquist rekonstruiert die aufgezeichnete Signalform perfekt, unabhängig von der Bitbreite und Samplingfrequenz.

Die Genauigkeit des Verfahrens und damit die Qualität der Wiedergabe hängt EINZIG und ALLEIN von folgenden Faktoren ab:

- exakte Zeitpunkte der Entnahme der Samples
- exakte Zeitpunkte beim Auslesen der Samples
- perfekte Filterung der Signalanteile oberhalb der aufzuzeichnenden / wiederzugebenden Bandbreite

Klangliche Auswirkungen

Oft wird von besserem Klang oder höherer subjektiver Auflösung bei erhöhten Samplingraten wie 96 oder 192 kHz berichtet.
Wie wir gesehen haben, steigt die Genauigkeit des rekonstruierten Signals mit höherer Samplerate nicht.

Eine höhrere Samplerate könnte genutzt werden, die Bandbreite zu erhöhen. Nach Nyquist ließe sich mit einer Samplingrate von 192 kHz die Bandreite auf 95 kHz erhöhen.

In der Praxis der Aufnahmestudios wird die Bandbreite auch bei 24 Bit/192 kHz gegenüber 16 Bit/44,1 kHz beibehalten.

Aus guten Gründen, denn weder Musikinstrumente noch Mikrofone, Lautsprecher und das Gehör des Menschen reichen höher als 20 kHz.

Woraus resultieren also die Klangunterschiede?

1. Bei 192 kHz kann ein weniger steilflankiges Anlogfilter verwendet werden. Entgegen der Ansicht der meisten vertrete ich die Ansicht, dass die Rolle der Filterung ohne jeden Nachteil das menschliche Gehör bilden kann.

2. Elektronische Filter können als Analog- oder Digitalfilter realisiert werden. Analoge Filter mit der geforderten sehr hohen Steilheit verursachen nicht akzeptable Phasendrehungen des Signals sowie sehr schlechte Impulseigenschaften. Digitalfilter (Oversampling) verhieben die Aliase in höhere Freauenzbereiche, so dass die analoge Filterung sanfter erfolgen kann, mit Flankensteilheiten von „nur noch“ 18 oder 24dB“. Dies sind immer noch recht steile Filter mit nicht wirklich optimalem Phasen- und Impulsverhalten. Zu den Nachteilen des Analogfilters kommen bei Verwendung von Digitalfiltern weitere Nebenwirkungen hinzu, auf die hier nicht weiter eingegangen werden soll.

3. Das beste Filter ist das menschliche Ohr. bei 22 kHz beträgt die Dämpfung bereits >140 dB, mehr als jedes Analog- oder Digitalfilter oder eine Kombination aus beidem erreichen. Dieses Filter ist beim menschlichen Hören sowieso immer in Betrieb, wir können es nicht entfernen und es wirkt ohnehin zusätzlich zu elektronischen Filtern. Das Problem ist, dass wir hinter dem Filter „Ohr" kein Messgerät anschließen können, das messtechnisch belegt, dass danach der originale Signalformverlauf im Gehirn ankommt. Davon darf allerdings ausgegangen werden aufgrund der Tatsache, dass jeder andere Filter, der die geforderten Dämpfungswerte sicherstellt, funktioniert.

Weniger steile Filter bei 192 kHz mit weniger Nebenwirkungen sind EIN Grund, warum 24 Bit/192 kHz DACs besser klingen *können* als 16 Bit / 44.1 kHz DACs. Beachte den Konjunktiv, denn die mögliche sanftere Filterung ist der einzige Vorteil einer höheren Samplerate. Dieser Nachteil kommt nur zum tragen, wenn tatsächlich ein elekronisches Filter zum Einsatz kommt, eine Maßnahme, die m.E. durch den Filter „Ohr" ersetzt werden kann.

Dem stehen eine Reihe von Nachteilen gegenüber:

- Kein 24 Bit/192 kHz DAC auf dem Markt erfüllt die Monotoniebedingung. Alle 24 Bit DACs haben 2-4 LSB Monotoniefehler, was unmittelbar an ihren Dynamikmesswerten ablesbar ist. Statt der theoretischen 144 dB erreichen die 24 Bit DACs nur 120-132 dB. Es ist ein Fehler, anzunehmen, dass dies in der Praxis keine Rolle spiele, weil auf der CD nur 96 dB aufgezeichnet sind.

- 4 LSB Biterror führen zu einem Dynamikfehler von 24 dB!!! Alle 24 Bit DACs komprimieren das Signal. Dynamikkompression erhöht den Pegel leiser Passagen und reduziert den Pegel lauter Passagen. Hierdurch bedingt werden leise Details der Musik besser hörbar, was vordergründig beurteilt mehr Auflösung vortäuscht. Der Preis für diese *vorgetäuschte* Auflösung ist dramatisch reduzierte Dynamilk, 24 dB sind kein Pappenstiel! 24 dB beziehen sich dabei nicht auf den Maximalwert des DACs (120 -24 = 96 dB), sondern auf die tatsächliche Dynamik des aufgezeichneten Signals. Von z.B. 50 dB (typ.- Jazz-Wert) bleiben gerade noch 26 dB übrig!

- Die Anforderungen an die Clockgenauigkeit, also die genauen Samplezeitpunkte, steigen bei einer Samplerate von 192 kHz gegenüber eine Samplerate von 44,1 kHz dramatisch an. Während die Clockgenauigkeit für 44,1 kHz/16 Bit mit 110 ps (Pikosekunden = 10^{-9} Sek.) noch problemlos mit einem einfachen Quartz sicherzustellen ist, steigt die gerforderte Timingpräzision bei 24 Bit/192 kHz um den Faktor 2^8 = 1024 (die Bitdifferenz 24-16) auf 0,1 ps. Diese Genauigkeit ist auch mit großem Aufwand nicht sicherzustellen, die maximal mögliche Genauigkeit beträgt 1 ps, also eine Zehnerpotenz ungenauer als gefordert! Timing-Jitter ist unvermeidlich. Jitter führt zu ungenauer Signalrekonstruktion, also Verzerrungen.

Das klangliche Resultat:

24 Bit-DACs verringern die Dynamik der Musik, dadurch werden leise Details lauter hörbar, was viele Hörer mit mehr Auflösung verwechseln.

Timing-Jitter führt zu Signalungenauigkeiten des ausgegebenen Analogsignals. Ungenauigkeiten machen sich stärker bei hohen als bei niedrigen Frequenzen bemerkbar. Die Hochtonwiedergabe wirkt diffus, S-Laute können zischeln und der gesamten Wiedergabe mangelt es an Energie, weil die Energie, wenn man so will, zeitlich verteilt wird.

Links

- Compact Disc Digital Audio
- Die Compact Disc
- Nyquist-Shannon-Theorem
- Samplingrate/Abtastrate
- Quantisierung
- Jitter in digitalen Audiosystemen
- Red Book