

Bessere digitale Audioqualität

KEITH HOWARD UNTERSUCHT DEN HINTERGRUND UND DIE PHILOSOPHIE HINTER ROB WATTS' ULTRALANGEN FILTERN, DIE IN CHORD ELECTRONICS DACS, INSBESONDERE DEM M SCALER, ZU FINDEN SIND

Is ist ein interessanter historischer Zufall, dass 1948 sowohl das Jahr war, in dem Columbia Records die LP ankündigte, als auch das Jahr in dem Claude Shannons berühmte Arbeit *A Mathematical Theory of Communication* [1] veröffentlicht wurde, in der die mathematische Grundlage der Signalabtastung und damit der digitalen Audiotechnik beschrieben wird. Es wird heute anerkannt, dass Shannon nicht der erste war, der sich auf dieses Gebiet begab; andere, darunter der englische Mathematiker Edmund Whittaker, hatten die Theorie ganz oder teilweise früher. Shannons Aufsatz war jedoch der eigentliche Startschuss für die Informationstheorie und ebnete - in Erwartung der notwendigen technischen Fortschritte - den Weg für die Digitalisierung aller Arten von kontinuierlichen Signalen, nicht nur von Audiowellenformen.

Nur wenige Audiophile wussten von diesem Bereich der Mathematik oder seiner möglichen Anwendung auf Audiosignale, bis die BBC 1972 begann, die Audiofestverbindungen zu ihren Sendern durch digitale 13-Bit-Nicam-PCM-Audioverbindungen mit einer Abtastrate von 32 kHz und Companding (Komprimierung für die Übertragung und anschließende Erweiterung beim Empfang) zu ersetzen. Die erste, vom Broadcasting House zum Sender Wrotham in North Kent, nahm am 14. September desselben Jahres den Betrieb auf, und das Netz wurde dann schrittweise ausgebaut. Später in den 1970er Jahren wurden die ersten digitalen Aufnahmegeräte entwickelt, die eine ausreichende Qualität aufwiesen, um von Pionierfirmen wie Denon und Decca verwendet zu werden. 1982/3 kam dann die Compact Disc auf den Markt - der erste digitale Tonträger für Musik.

Angeichts der Tatsache, dass es 35 Jahre her ist, dass die CD in Europa auf den Markt kam, könnte man annehmen, dass die meisten Audiophilen die Grundlagen der Sampling-Theorie inzwischen fest im Griff haben, aber in Wahrheit wird sie

immer noch weitgehend missverstanden. Und zwar nicht nur von Audio-Amateuren, sondern auch von Audio-Profis, wenn Rob Watts, Chord Electronics' Digital Design Consultant, hat Recht. Seit Jahren wehrt sich Watts gegen die in der Audioindustrie übliche Praxis, relativ kurze - manchmal sogar sehr kurze - Digitalfilter in DACs mit hoher Überabtastung zu verwenden. Mit den technischen Möglichkeiten wurden seine Filter immer länger. Und jedes Mal, wenn die Filterlänge erhöht wird, verbessert sich seiner Meinung nach die Klangqualität. Watts' langes Bemühen um eine ausreichend lange

Die Erkenntnis, dass eine weitere Erhöhung des Filters nicht von subjektivem Nutzen ist, fand ihre jüngste Bestätigung in der Ankündigung des Chord Electronics *M Scaler* (siehe Kasten) auf der Londoner CamJam-Messe im Juli, einem Upsampler mit digitalem Eingang und digitalem Ausgang, der zum ersten Mal über eine Million Filtertaps verfügt, *d. h.* das verwendete linearphasige FIR-Interpolationsfilter (Finite Impulse Response).

hat mehr als eine Million Koeffizienten: 1.015.808, um genau zu sein. Um diese Zahl ins rechte Licht zu rücken,

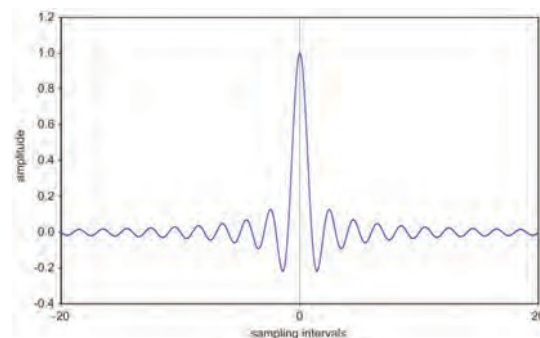
Die meisten Oversampling-DACs verwenden Filter, die höchstens ein paar hundert Taps (Koeffizienten) lang sind. Sogar *Dave*, der beste aktuelle Standalone-DAC von Chord, verwendet "nur" 164.000 Filtertaps.

Um Watts' unermüdliches Streben nach längeren Interpolationsfiltern zu verstehen - wobei jeder Schritt eine noch größere Kluft zwischen ihm und der anerkannten Industriepraxis aufgerissen hat -, muss man zu den Wurzeln zurückgehen: zur Shannon'schen Abtasttheorie und insbesondere zur "sinc"-Funktion.

Abb. 1. Zentraler Teil der Funktion $\text{sinc}(x)$

$\text{sinc}(x)$ ist die mathematische Abkürzung für die Funktion $\sin(x)/x$, die (in ihrem mittleren Teil) wie in Abb. 1 aussieht. Der Grund dafür, dass die x-Achse (horizontale Achse) mit "Abtastintervalle" beschriftet ist, ist wird in Kürze deutlich werden. Shannon hat in seiner berühmten Arbeit gezeigt, dass jedes bandbegrenzte kontinuierliche Signal - d. h. jedes analoge Signal mit einer strikten Begrenzung seiner Maximalfrequenz - exakt als eine Summe von zeitlich beabstandeten $\text{sinc}(x)$ -Wellenformen beschrieben werden kann.

Und nicht nur das: Wenn das Signal abgetastet wird, *d. h.* wenn seine Amplitude in regelmäßigen Abständen mit einer Rate gemessen wird, die



mindestens doppelt so hoch wie die höchste Signalfrequenz ist, stellt jede Abtastamplitude die Amplitude der zugehörigen $\text{sinc}(x)$ -Wellenform dar, die auf diesen Abtastpunkt zentriert ist. An allen anderen Abtastpunkten ist der Wert dieser speziellen $\text{sinc}(x)$ -Funktion gleich Null, genauso wie der Wert der $\text{sinc}(x)$ -Funktionen, die auf jeden anderen Abtastpunkt zentriert sind, hier gleich Null ist. Durch die beschriebene Abtastung der Wellenform werden also alle Informationen extrahiert, die zur Rekonstruktion der Wellenform erforderlich sind, und zwar *mit absoluter Genauigkeit*.

Dies ist der Analog-Digital-Wandlungsprozess, und er ist durchaus praktikabel. Alles, was wir hinzufügen müssen, um es in der Praxis zu realisieren, ist die Quantisierung der Amplitudenmessung, so dass jeder Abtastwert durch eine Zahl von endlicher Länge dargestellt werden kann.

Im Gegensatz dazu ist der Prozess der Wellenformrekonstruktion - die Digital-Analog-Wandlung - nicht so einfach. In einer idealen Welt würde dies durch die Erzeugung eines Impulses mit geeigneter Amplitude für jeden Abtastpunkt und die Weitergabe der Folge von regelmäßig aufeinanderfolgende Impulse durch einen idealen Tiefpassfilter, dessen oberer Rand des Durchlassbereichs auf die Hälfte der Abtastrate eingestellt ist. Die Impulsantwort eines solchen Filters ist die Sinc-Funktion, so dass jeder Impuls eine Sinc-Wellenform der erforderlichen Amplitude erzeugen würde und die Folge von Sinc-Funktionen die ursprüngliche Wellenform wiederherstellen würde.

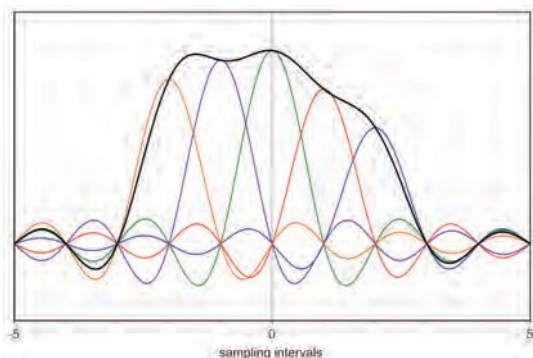


Abbildung 2. Wie die Summierung von $\text{sinc}(x)$ -Funktionen - eine pro Signalabtastung - die Wellenform zwischen den Abtastpunkten aufbaut

Dies wird in Abb. 2 veranschaulicht, die sechs aufeinanderfolgende Sinusfunktionen mit unterschiedlicher Amplitude und ihre Summe (die schwarze Kurve) zeigt. Jede sinusförmige Funktion trägt nichts zu der summierten Signalamplitude an anderen Abtastpunkten bei, wohl aber zur Wellenform *zwischen den Abtastpunkten*. Es ist ein

häufiges Missverständnis der Abtastung anzunehmen, dass die Wellenform zwischen den Abtastpunkten nicht bekannt ist, aber das ist nicht wahr - vorausgesetzt, das Eingangssignal ist bandbegrenzt und wird mindestens doppelt so schnell abgetastet als ihre höchste Frequenzkomponente, wie es das Shannon-Abtastverfahren erfordert. In diesem Fall kann die Wellenform zwischen den Abtastpunkten sein

Die Wellenform hängt nicht nur vom Wert der nahegelegenen Abtastwerte ab, sondern letztlich auch vom Muster der Abtastwerte im gesamten abgetasteten Signal.

In der Praxis ist dieses theoretische DAC-Schema aus zwei Gründen nicht realisierbar. Erstens gibt es den idealen Tiefpassfilter mit einer unendlichen Abschwächung am Rand des Durchlassbereichs nur in Abstraktionen - in der realen Welt, in der Filter immer endliche Abschwächungsraten haben, kann er nur angenähert werden. Zweitens: Selbst wenn der perfekte Tiefpassfilter kein Traum wäre, würde dieser Ansatz wegen der geringen Energiemenge, eindeutig rekonstruiert werden. Wichtig ist, dass die

die in jedem Impuls enthalten ist, einen unzureichenden Signal-Rausch-Abstand liefern.

Bei realen DACs müssen zwei Kompromisse eingegangen werden. Erstens wird die Amplitude jeder Abtastung nicht als Impuls dargestellt, sondern als Schritt, der für eine ganze Abtastperiode beibehalten wird. Dieser "Sample-and-Hold"-Prozess vermeidet die Signal-Rauschabstandsproblem, sondern führt zu einem nicht flachen Frequenzgang, der zur Nyquist-Frequenz (halbe Abtastrate) hin leicht abfällt. Die Lösung ist trivial: Der Abfall kann durch eine Entzerrung korrigiert werden, was in der Regel auch geschieht. Den zweiten Kompromiss habe ich bereits angedeutet. Da ein idealer Tiefpassfilter nicht möglich ist, muss ein Filter mit einer langsameren Abschwächung eingesetzt werden - und das liefert dann keine $\text{sinc}(x)$ -Impulsantwort.

Dieser letzte Punkt wird üblicherweise einfach ignoriert. Der analoge Ausgangsfilter (oder bei überabgetasteten Systemen der digitale Interpolationsfilter) gilt als angemessen, wenn er eine ausreichend gute Leistung im Frequenzbereich erzielt, d. h. einen ausreichend flachen Durchlassbereich und eine angemessene Dämpfung der Bildfrequenzen, die oberhalb der Nyquist-Frequenz auftreten. Abbildung 3 zeigt den Frequenzgang von ein Beispiel für einen Interpolationsfilter, der (unter Verwendung der bekannten Parks-McClellan-Äquiripple-Methode) für die 4fache Überabtastung von 44,1kHz-Daten mit den folgenden Spezifikationen entwickelt wurde:

Obere Frequenz des Durchlassbereichs 20kHz
Durchlassbereichswelligkeit 0,01dB
Sperrbereich untere Frequenz 24kHz

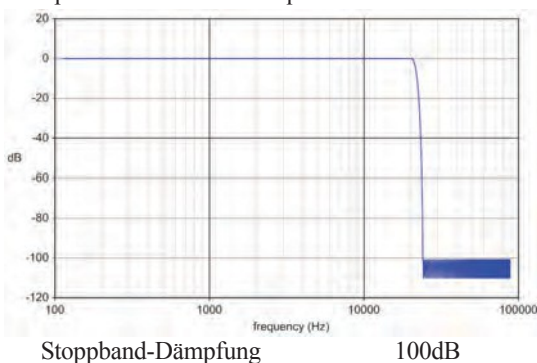


Abb. 3. Frequenzgang eines beispielhaften 4x-Interpolationsfilters

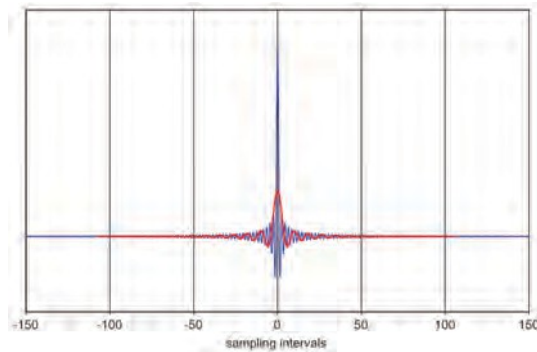


Abb. 4. Impulsantwort der Interpolationsfilters aus Abb. 3 (rote Kurve), überlagert mit der Funktion $\text{sinc}(x)$ (blaue Kurve)

Der resultierende FIR-Filter hat 215 Koeffizienten (Abgriffe) und, da er linearphasig ist, eine zeitsymmetrische Impulsantwort. Die Dämpfung bei der Nyquist-Frequenz (22,05kHz) beträgt etwa 10dB. Die Impulsantwort des Filters ist in Abb. 4 dargestellt, wobei sie die Sinusfunktion aus Abb. 1 überlagert, hier jedoch über einen größeren Bereich von Abtastintervallen, um die Anzahl der Filterkoeffizienten zu berücksichtigen. In Abb. 5 sind die Daten von Abb. 4 wiederholt, jedoch auf einer Dezibel-Amplitudenskala dargestellt. Aus beiden Diagrammen geht klar hervor, dass der Interpolationsfilter - obwohl er die repräsentativen Frequenzbereichskriterien erfüllt - eine Impulsantwort hat, die sich von der $\text{sinc}(x)$ -Funktion deutlich unterscheidet, und das nicht nur, weil sie kürzer ist.

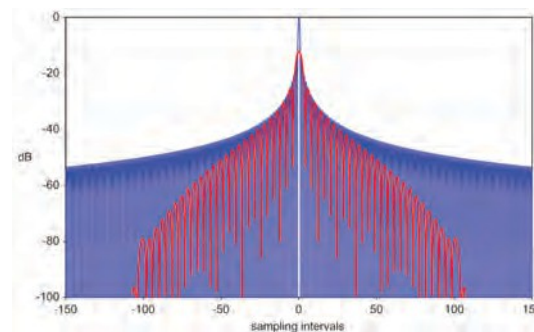


Abb. 5. Eine Wiederholung von Abb. 4, aber diesmal mit Dezibel-Amplitudenskala

Abb. 5 verdeutlicht einen wichtigen Punkt: Die Hüllkurve der $\text{sinc}(x)$ -Funktion - die für Werte von x zwischen minus unendlich und plus unendlich endlich ist - nimmt mit der Zeit langsam ab. Bei 150 Abtastintervallen von ihrem zentralen Spitzenwert ist die Hüllkurve nur um etwas mehr als 50 dB abgeklungen. Die naheliegende Frage lautet: Um wie viel muss sie abklingen, damit ihr Beitrag zur Wellenform zwischen den Abtastintervallen unbedeutend wird? Diese Frage ist nicht einfach zu beantworten, aber wenn wir sagen, dass die Hüllkurvenamplitude um 100 dB

unter die 16-Bit-Rauschgrenze für ein 0 dBFS (full scale) Sample fallen muss, können wir leicht berechnen, welcher Ausschnitt der Sync-Funktion erforderlich ist. Die Hüllkurve der Sync-Funktion wird allein durch den Nenner bestimmt

von $\sin(x)/x$, verhält sich also wie $1/x$ mit $x = N\pi$ (Winkel im Bogenmaß), wobei N die Anzahl der Abtastintervalle im Abstand von der zentralen Spitze ist. Damit die Hüllkurve 100 dB unter ihrem Spitzenwert liegt, ist $1/x = 0,000001$, was $N = 31,831$ entspricht. Diese Abtastlänge ist auf beiden Seiten der zentralen Spitze erforderlich, so dass die Gesamtlänge des $\text{sinc}(x)$ -Auszugs doppelt so lang ist. Mit anderen Worten: Bei einer Abtastrate von 44,1 kHz beträgt die Gesamtlänge des erforderlichen sinc -Funktionsausschnitts 1,443 Sekunden - das entspricht ziemlich genau der Filterlänge, die der *M-Scaler* von Chord bietet.

Dies erklärt, kurz gesagt, Watts' Streben nach noch nie dagewesenen langen Interpolationsfiltern, die er mit der so genannten Windowed-Sinc-Technik entwickelt. Wie im vorigen Absatz bereits angedeutet, wird dabei ein Stück aus der Mitte der Sinusfunktion extrahiert, aber für optimale Ergebnisse wird dieser Prozess subtiler sein als ein einfaches "Anheben" der Werte der sinc -Funktion und Abschneiden des Rests. Bessere Ergebnisse werden erzielt, wenn die exzerpierte sinc -Funktion

ist gefensterter, *d. h.* geformt, um plötzliche Abbrüche an beiden Enden zu vermeiden. Watt's WTA (Watts Time Alignment) Fensterungsalgorithmus ist ein streng gehütetes Geheimnis, und

Er musste mit zunehmender Filterlänge verfeinert werden, aber sein Name weist auf Watts' wichtigstes Entwurfskriterium hin: die Beibehaltung eines präzisen Einschwingverhaltens.

Soweit mir bekannt ist, ist kein anderer Designer in Watts' Fußstapfen getreten. Und Sie müssen annehmen, dass es wäre eine gewaltige Aufgabe, wenn man bedenkt, dass Watts seinen einsamen Weg schon seit Jahrzehnten beschreitet. Aber wenn es jemand versuchen wollte - vor allem angesichts der allgemein positiven Resonanz auf die digitalen Produkte von Chord - wie könnte er es anstellen?

Als Erstes müssen Sie sich davon überzeugen, dass Watts' Ansatz richtig ist. Das geht natürlich ganz einfach, indem man sich die Produkte von Chord anhört, idealerweise eine Auswahl von ihnen, die den Weg zu einer größeren Filterlänge aufzeigen. Aber es gibt noch einen anderen Weg, und zwar einen. Das ist viel billiger als die Anschaffung einer Sammlung von Akkord-Hardware und viel einfacher als der große Schritt der Programmierung eines FPGA. Das heißt, die Sinc-Interpolation offline, in Software, durchzuführen.

Ich habe vor über 10 Jahren ein Softwareprogramm geschrieben, das dies ermöglicht - und im Vergleich zur FPGA-Programmierung ist es ein absolutes Kinderspiel. Das Problem ist, dass das Programm ewig braucht, um mit etwas Längerem als einer sehr kurzen Audiodatei zu laufen, weil es die Berechnung von $(U-1)$

$\times N^2 \sin(x)/x$ Werte, wobei U der Überabtastungsfaktor und N die Anzahl der Abtastwerte in der Datei ist. Dies gilt für jeden Kanal. Aber, wie gesagt, es ist einfach, es ist billig, und es erlaubt Ihnen, eine Datei zu erzeugen und anzuhören, die mit voller Sinc-Interpolation überabgetastet wurde, in dieser Hinsicht ist es sogar besser als der *M-Skalierer*.

Sobald Sie die Datei haben, können Sie sie als Referenz verwenden, um andere Dateien zu prüfen, die mit

Interpolationsfilter mit endlicher Länge in verschiedenen Ausführungen. Um Ihnen ein Beispiel zu zeigen, habe ich den Code mit einer kurzen (0,98 Sekunden) Mono-WAV-Datei mit 44,1 kHz/16 Bit ausgeführt, die eine einzelne auf einem Cembalo gespielte Note enthält. Für 4× Oversampling, die Verarbeitung (die 64-Bit-Gleitkomma-Arithmetik verwendet und eine 24-Bit-WAV-Ausgangsdatei erzeugt) dauerte 295 Sekunden - über 300× Echtzeit - die auf einem einzigen Prozessorkern meines alternden Desktop-Computers läuft.

In Abb. 6 ist das Spektrum der Originaldatei (rote Kurve) und das der überabgetasteten Datei (blaue Kurve) überlagert. Es zeigt (a), dass sich die beiden über den Durchlassbereich überlappen, wie es sein sollte, und (b), dass die Sinc-Interpolation tatsächlich zu einer Tiefpassfilterung bei 22,05 kHz führt, wobei das Grundrauschen oberhalb dieser Frequenz auf Dither zurückzuführen ist. (Die Originaldatei war mit einer 4096-Punkte-FFT und die interpolierte Datei mit einer 16.384-Punkte-FFT analysiert, um sicherzustellen, dass die Spektren die gleiche Frequenzauflösung haben).

Ich habe in der Vergangenheit, wenn auch noch nie in gedruckter Form, vorgeschlagen, dass jemand mit Zugang zu schwergewichtigen

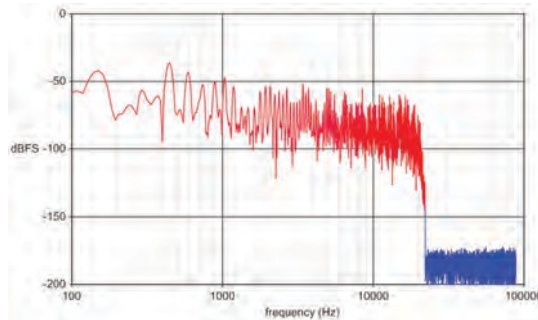


Abb. 6. Spektren einer kurzen Aufnahme einer einzelnen Note, die auf einem Harfenakkord gespielt wird. Die rote Kurve zeigt die 44,1kHz/16-Bit-Originaldatei, die blaue Kurve eine 4fach überabgetastete Version, die mit voller Sinusinterpolation erzeugt wurde

Zahlenjongleure könnten damit einen leicht zugänglichen Zwischenspeicher für sinc-interpolierte Musikdateien anlegen, um bessere Interpolationsfilter zu entwickeln. Das ist vielleicht kein Akt der Philanthropie, der Ihnen ein Titelblatt des Time Magazine sichern würde - aber Audiophile könnten Ihren Namen für immer verehren. Hat jemand Interesse?

Referenz

1) Shannon, C. E. "A Mathematical Theory of Communication", *Bell System Technical Journal* (1948). Nachgedruckt in Buchform, mit einer Einführung für Laien von Warren Weaver, als "The Mathematical Theory of Communication", University of Illinois Press.

Chord Electronics M Scaler

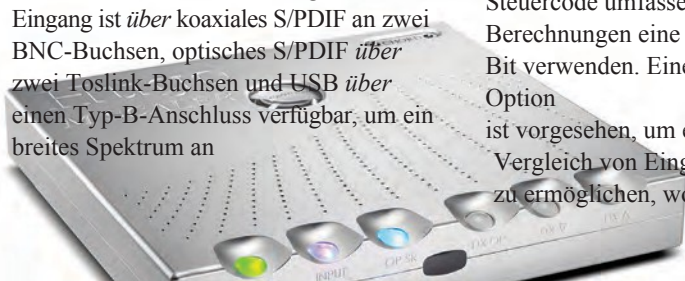
Ich habe das Gefühl, dass sich der *M Scaler* als eines der bedeutendsten und wahrscheinlich kontroversesten Produkte des Jahres 2018 erweisen wird. Er wird mit Sicherheit diejenigen faszinieren, die Rob Watts' DAC-Designs bereits für und ebenso wahrscheinlich die Monty Montgomerys dieser Welt veranlassen, sie als wahnhaft abzutun.

Der *M Scaler* nutzt die Oversampling-Technologie des Chord *Blu MkII* CD-Transporters und verpackt sie in ein Gehäuse mit den Maßen 40,5×235×236 mm (hwd), 2,55 kg schweres Gehäuse zu weniger als der Hälfte des Preises. Der digitale Eingang ist über koaxiales S/PDIF an zwei BNC-Buchsen, optisches S/PDIF über zwei Toslink-Buchsen und USB über einen Typ-B-Anschluss verfügbar, um ein breites Spektrum an

von digitalen Quellen. Die Ausgabe erfolgt über einen einzelnen BNC-Anschluss mit 352,8/384kHz, über einen optischen S/PDIF-Anschluss mit 176,4/192kHz oder, für volle Leistung in Verbindung mit dem *Qutest*, über *Hugo TT2* oder *Dave*, über zwei BNC-Buchsen bei 705,6/768kHz.

Die Schlüsselkomponente des *M-Scalers* ist der Xilinx *XC7A200T* FPGA-Chip (Field Programmable Gate Array), der 740 DSP-Kerne bereitstellt. Die neueste Filterarchitektur von Watts (die auch im *Hugo TT2* zum Einsatz kommt) verwendet 528 dieser Kerne, die mit einer 4096-fachen Abtastfrequenz arbeiten, eine halbe Million Zeilen Steuercode umfassen und bei den Berechnungen eine Auflösung von 56 Bit verwenden. Eine Pass-Through-Option ist vorgesehen, um einen sofortigen Vergleich von Eingabe und Ausgabe zu ermöglichen, wobei

Verstärkungskorrektur, um sicherzustellen, dass es keine Pegelunterschiede gibt. [Oversampling kann zu interpolierten Abtastwerte, die 0 dBFS (Vollausschlag) überschreiten, was eine Gewinnminderung, um sie unterzubringen]. Erste Tests mit 512.000 Zapfstellen



ergab, was Watts als "völlig unerwartete" Verbesserung gegenüber den 164.000 Anzapfungen des *Dave* bezeichnet, so dass das Ziel auf über eine Million Anzapfungen angehoben wurde. Bei der Ausgangsbitrate von 705,6/768 kHz ($16 \times 44,1/48$ kHz) bedeutet dies eine Latenz - die Verzögerung zwischen Eingabe und Ausgabe - von etwa 0,6 Sekunden, während der *M-Scaler* seine Berechnungen mit dem ersten Sample durchführt. Da diese Verzögerung zu inakzeptablen Lippensynchronisationsproblemen führen würde, wenn der Ton das Video begleitet, bietet der *M-Scaler* einen Videoeingang, der einen asymmetrischen Interpolationsfilter, der die Latenzzeit auf akzeptable 0,1 Sekunden reduziert.

Was sind die wahrgenommenen Vorteile für die Klangqualität? Watts sagt, dass die verbesserte Transientengenauigkeit des längeren Filters die Klangfarbe von Instrumenten klarer macht, den Bass strafft und die Klangbühne "dramatisch" öffnet. "Nachdem man den *M Scaler* gehört hat", sagt er, "ist es sehr schwierig, den *Hugo TT2* oder *Dave* zu hören." Und ist sein Durst nach mehr Filterabgriffen nun gestillt? Nein: "Mein Bauchgefühl sagt mir, dass wir noch weiter gehen müssen."

Der M Scaler wird im Herbst zu einem Preis von 3495 £ erhältlich sein.