

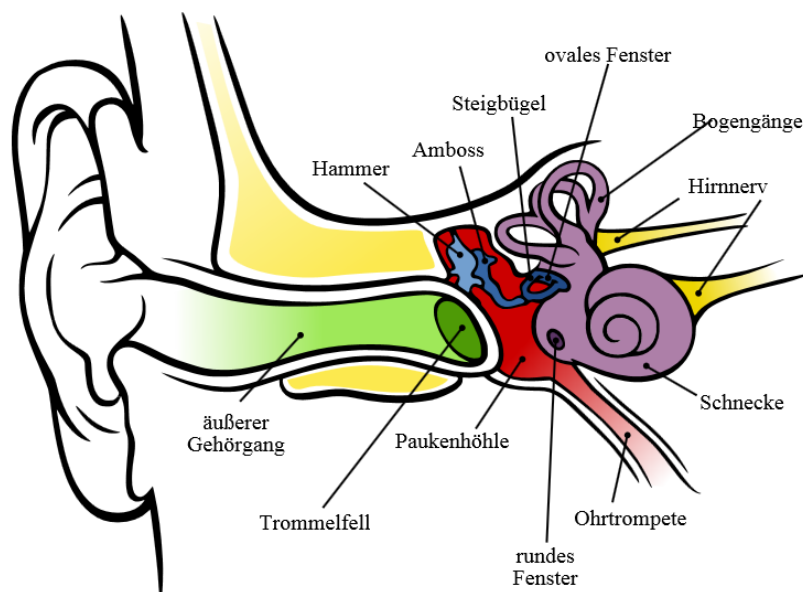
Sinn und Unsinn von hohen Samplingraten

Die Entwicklung der Computer- und Digitaltechnologie ermöglicht immer höhere Sampleraten bei Analog-zu-Digital- und Digital-zu-Analog-Wandlung. War zu Beginn der Digitalära in den 70er Jahren des letzten Jahrhunderts eine Abtastrate von 48kHz und damit ein nutzbarer Audiobereich bis 22kHz das höchste der Gefühle, sind es inzwischen 384kHz bei PCM (entsprechend einem nutzbaren Frequenzbereich von 190kHz) bzw. bei DSD 5,64MHz (entsprechend einem nutzbaren Frequenzbereich von ca. 50kHz). Zum einen möchten die Hersteller entsprechenden Equipments natürlich mit diesen Weiterentwicklungen immer neue Geräte verkaufen. Gleichzeitig gibt es auch eine breite Strömung innerhalb der High-End- und Audiophilen-Community, die diese Entwicklung zu höheren Abtastraten als beständige Verbesserung der Klangqualität begrüßt. Allerdings gibt es gute Gründe zu hinterfragen, ob dieses beständige Höher, Weiter, Mehr wirklich der Qualität des reproduzierten Klanges dient.

Die Frage nach der sinnvollen Samplingrate läuft auf die Frage nach dem sinnvollerweise für die Audio-Reproduktion verfügbaren Frequenzbereich hinaus. Um diese Frage näher zu beleuchten wollen wir im folgenden zuerst das menschliche Hörvermögen und anschließend die zu reproduzierenden Schallereignisse näher anschauen.

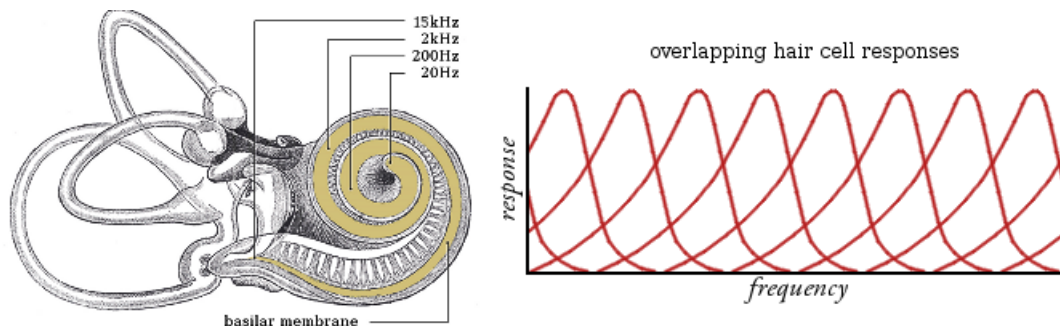
I. Das menschliche Gehör

Wenn es um das menschliche Gehör geht, sollte man mit den Ohren beginnen. Das menschliche Gehör ist ein Sinnesorgan von beeindruckender Präzision. Unser Ohr, beziehungsweise das gesamte auditive System, verwandelt Schallwellen in elektrochemische Impulse, die vom Nervensystem an das Gehirn übermittelt werden, wo weitere Prozesse zur Bearbeitung und Evaluierung des Signals stattfinden. Die physiologischen und neuronalen Begebenheiten unseres Gehörs ermöglichen es uns, daraus auf Parameter wie Tonhöhe oder Lautstärke sowie auf Entfernungen und die Richtung, aus der das akustische Signal in Bezug auf den Hörer kommt, zu schließen – und das bei gleichzeitiger Wahrnehmung vieler einzelner, sich überlagernder Signale.



Graphik 1: Das menschliche Gehör

Die auf das Ohr treffenden Schallwellen werden von der Ohrmuschel aufgenommen und durch den äußeren Gehörgang zum Trommelfell geleitet. Entsprechend der Intensität sowie der Frequenz der Schallwellen wird das Trommelfell in Schwingungen versetzt. Die Schwingungen des Trommelfells werden wiederum auf die Gehörknöchelchenkette (Steigbügel, Hammer und Amboss) übertragen und erreichen so das ovale Fenster. Durch die Gehörknöchelchen wird der Schall zudem um das 20fache verstärkt. Träfe der Schall direkt auf das ovale Fenster, würden nur etwa zwei Prozent der Schallenergie aufgenommen werden. Der Weg des Schalls wird bis zu diesem Punkt auch als Luftleitung bezeichnet. Die Schwingungen des Steigbügels am ovalen Fenster versetzen nun die lymphähnliche Flüssigkeit (Perilymphe) der sog. Schnecke in Schwingungen und durchlaufen sie als sogenannte Wanderwellen. Dadurch werden die feinen Haarsinneszellen verbogen und es entsteht ein Aktionspotential, welches an die Nervenfasern übertragen wird. Von dort aus ziehen die Fasern zum Gehirn (Hörzentrum), wo die ankommenden Potentiale verarbeitet und als Ton interpretiert werden. Die Umwandlung von Tonsignalen in Nervenimpulse und somit Wahrnehmung erfolgt somit in der Schnecke (Cochlea):



Graphik 2: Sinneswahrnehmung in der Cochlea

Das menschliche Gehör hört durch vier Reihen von Härchen, die sich in der Basilarmembraninnerhalb der Cochlea befinden. Jede Haarzelle ist auf ein schmales Frequenzspektrum, das durch seine Position auf der Membran bestimmt wird, gestimmt. Da die Basilarmembran am ovalen Fenster schmal und dick ist, ist ihre Eigenfrequenz hier hoch (bei geringer Amplitude), so dass in diesem Bereich hohe Frequenzen wahrgenommen werden. Weiter in Richtung Zentrum der Schnecke (Helicotrema), wo sie breiter und dünner ist, schwingt sie mit niedrigerer Frequenz (bei größerer Amplitude), so dass in diesem Bereich tiefere Frequenzen wahrgenommen werden. Die Empfindlichkeit der jeweiligen Haarzelle hat ihr Maximum in der Mitte des Frequenzspektrums und fällt rasch zu beiden Seiten hin ab, so dass sich die Spektren nebeneinander liegender Härchen überlappen und eine Art überlappenden Serie von Frequenzfiltern darstellt.

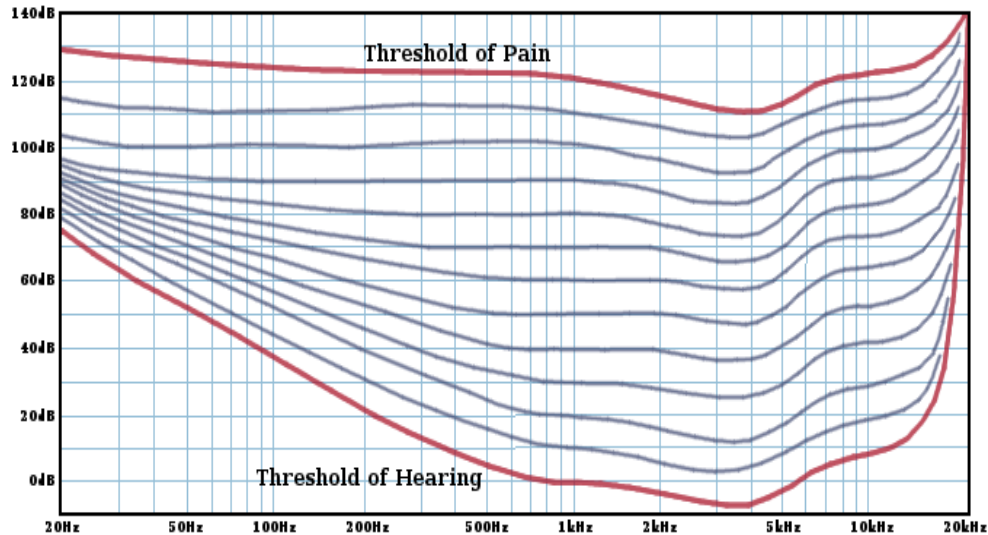
Schallereignisse werden physikalisch in der Hauptsache durch drei Größen charakterisiert:

1. Die Frequenz oder Tonhöhe des Tonsignals in Schwingungen pro Sekunde in Hertz (Hz) angegeben und
2. die Intensität oder der Schalldruckpegel des Tonsignals, zumeist in Dezibel (dB) angegeben.
3. Der Zeitliche Verlaufen von beiden

Ein Rückschluss von Schalldruckpegel auf die vom Menschen wahrgenommene Lautstärke ist nur sehr eingeschränkt möglich. Da der Schalldruckpegel nicht proportional bestimmt wird, sondern logarithmisch. Das heißt, ein Ton mit 60 dB ist nicht sechs mal lauter als ein Ton mit 10 dB, sondern besitzt die 600fache Intensität. Als Faustformel gilt, dass 10 dB Unterschied etwa als doppelte bzw. halbe Lautstärke wahrgenommen wird.

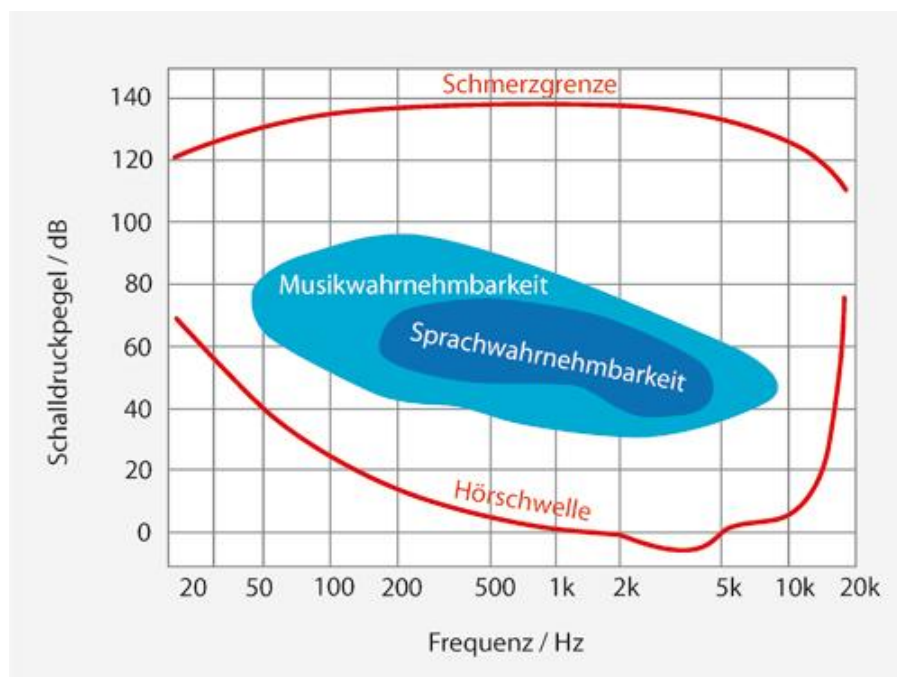
Das liegt daran, dass die vom Menschen wahrgenommene Lautstärke als psychoakustische Größe, von mehreren Faktoren abhängt: neben dem Schalldruckpegel vor allem auch vom Frequenzspektrum. Die Lautstärke, gemessen in Phon, ist ein psychoakustisches Vergleichsmaß, das beschreibt, welchen Schalldruckpegel ein Sinuston mit einer Frequenz von 1.000 Hz haben müsste, damit dieser Ton genauso laut empfunden wird, wie das betrachtete Hörereignis. Bei einer Schall-Frequenz von 1.000 Hz stimmen Schalldruckpegel, gemessen in Dezibel, und Lautstärkepegel, gemessen in Phon, überein. Für Sinustöne anderer Frequenzen sowie für komplexe Schallereignisse sind dagegen andere Schalldruckpegel erforderlich, um den

gleichen Lautstärkeindruck zu erzielen. Welcher Schalldruckpegel für einen Einzelton bei welcher Frequenz erforderlich ist, um jeweils den gleichen Lautstärkeindruck (in Phon) zu erzielen, ist durch eine Vielzahl an Experimenten im Laufe der letzten 100 Jahre für Menschen mit gesundem Hörvermögen ermittelt worden, die natürlich individuell schwanken können und in den „Kurven gleicher Lautstärkepegel“ (Isophone für den Bereich 20 - 100 Phon) gem. ISO 226:2003 definiert wurden:



Graphik 3: Isophone des menschlichen Gehörs

Auf jeder der blauen und roten Linien ist die wahrgenommene Lautstärke für den Bereich 0 - 120 Phon in etwa gleich. Zum Beispiel: Ein 20Hz Ton mit einem Pegel von 95 dB wird so laut empfunden, wie ein 4kHz Ton mit einem Pegel von 45 dB, nämlich beide mit einer Lautstärke 50 Phon. Um eine Annäherung des gemessenen Schalldruckpegels an die vom Menschen wahrgenommene Lautstärke zu erreichen, wird in der akustischen Messtechnik der gemessene Schalldruckpegel auch mit den Isophonen Funktionen bewertet und als "bewerteter Schalldruckpegel" in dB(A) angegeben. Aus den Isophonen ergibt sich die Hörfläche des Menschen:



Graphik 4: Hörfläche des Menschen

Das menschliche Gehör bzw. sein Hörbereich wird typischerweise nach

- unten begrenzt durch die Hörschwelle (auch Ruhehörschwelle). Das ist derjenige Schalldruckpegel, bei dem das menschliche Gehör Töne oder Geräusche gerade noch wahrnimmt. Sie liegt bei 1 kHz und einem Schalldruckpegel von null Dezibel (dB).
- nach oben begrenzt durch die akustische Schmerzschwelle, was derjenige Schalldruckpegel ist, bei dem das menschliche Gehör Töne oder Geräusche als sehr unangenehm wahrnimmt.

Sowohl bei tieferen Frequenzen als auch bei hohen ist die Hörschwelle zu höheren Pegeln hin verschoben. So können normalhörende Menschen einen tiefen Sinuston von 30 Hz als auch ein hoher von 15 kHz erst ab einem Schalldruckpegel von ungefähr 60 dB wahrnehmen. Am empfindlichsten ist das menschliche Gehör für Frequenzen zwischen 3500 und 4000 Hz. In diesem Frequenzbereich vermag es Schalldruckpegel von bis minus 5 dB zu registrieren. Das obere Limit des menschlichen Hörvermögens ist definiert, als der Punkt, an dem sich die Hörschwelle mit der Schmerzgrenze kreuzen (bei Kindern max. 20.000Hz - ansonsten bei Jugendlichen und Erwachsenen 16.000 bis 18.000Hz - und ca. 140 dB). Allerdings ist die höchste Frequenz, die wahrgenommen werden kann, vom Alter und Gesundheitszustand abhängig. Mit zunehmendem Alter steigt die Hörschwelle vor allem bei höheren Frequenzen deutlich an (Presbyakusis).

Darüber hinaus ist keine Gehörwahrnehmung möglich, egal wie laut das Signal auch sein mag. Niemand nimmt Töne ober- oder unterhalb dieser Grenzen wahr. Die in Audio- und High-End-Diskussionen immer wieder auftauchenden Behauptungen, dass Menschen durch Modulationen oder wie auch immer Töne über 20kHz wahrnehmen können, wird durch wissenschaftliche Analysen nicht bestätigt.

II. Digitalisierung analoger Tonsignale

Wie ich ausführlich im Artikel über "Digitalisierung von Audiodaten" erläutert habe, werden beim Digitalisieren die Toninformation auf zwei Kenngrößen hin moduliert:

- **Abtastrate:** auch Samplingrate, bezeichnet die Häufigkeit, mit der das kontinuierliche Tonsignal in einer vorgegebenen Zeit abgetastet und in ein zeitdiskretes Signal umgewandelt wird (Sampling). Sie ist somit ein Maß für die zeitliche Auflösung des Tonsignals und bestimmt den Frequenzumfang. Unter Einhaltung des **Nyquist-Shannon-Abtasttheorems** können Frequenzen bis zur halben Abtastrate verlustfrei kodiert und wieder rekonstruiert werden, so dass die aus den digitalen Daten rekonstruierte Wellenform exakt der ursprünglichen analogen Wellenform entspricht. Die Digitalisierung beeinflusst weder den Frequenzverlauf noch das Phasenverhalten. Allerdings setzt das Theorem ideale Antialias- und Rekonstruktionsfilter voraus. Somit ergeben sich folgende maximale Frequenzumfänge für übliche Abtastraten:

Abtastrate	Frequenzumfang
44,1kHz	1Hz - 22,1kHz
88,2kHz	1Hz - 44,1kHz
96,0kHz	1Hz - 48,0kHz
192,0kHz	1Hz - 96,0kHz

- **Wortbreite:** auch Bittiefe bezeichnet die Messpunkte zu jedem der Abtastpunkte (Quantisierung). Sie geben die Amplitude des kontinuierlichen Signals zu diesem Zeitpunkt in Bit an und legen somit die dynamische Auflösung der Digitalisierung fest. Die Wortbreite einer Digitalisierung bestimmt den

Dynamikumfang. Da 1 Bit ca. 6 dB(A) an Dynamikumfang entspricht, ergeben sich folgende maximale Dynamikwerte für übliche Wortbreiten:

Bittiefe	Max. Amplitudenwerte	Dynamikbereich
8 Bit	$2^8 = 256$	48 dB
16 Bit	$2^{16} = 65.536$	96 dB
24 Bit	$2^{24} = 16.777.216$	144 dB
32 Bit	$2^{32} = 4.294.967.296$	192 dB

Es sei an dieser Stelle angemerkt, dass ungeachtet der Werte der Tabelle, der effektive Dynamikumfang für jede Worttiefe durch Dithering und Noise Shaping erweitert werden kann. Dadurch ist es z. B. möglich, den effektiven Dynamikumfang von 16Bit Aufnahmen auf ca. 100 – 120dB zu erweitern. Umgekehrt sei auch erwähnt, dass echte 24Bit oder darüber hinaus mit heutigen Technologien kaum zu realisieren sind. Der Grund hierfür liegt im thermischen Rauschen der verwendeten elektronischen Bauteile, insbesondere der Widerstände, bei Raumtemperatur, das sich etwas oberhalb von -140 dB abspielt.

Im Gegensatz zum Sampling, ist die Quantisierung kein verlustfreier Prozess. Die Wahl der Worttiefe ist willkürlich. Einen kontinuierlichen Amplitudenverlauf kann ein quantisiertes Digitalsignal egal in welcher Worttiefe nicht verlustfrei reproduzieren.

Für das Beispiel des Red Book (CD) Standards heißt das, dass mit der dort verwendeten Abtastrate von 44,1 kHz theoretisch Tonsignale bis 22.050Hz verlustfrei digitalisiert werden können. Da das menschliche Gehör, maximal Töne von ca. 20Hz bis 20.000Hz wahrnehmen kann, sollte die Digitalisierung nach dem Red Book Standard zumindest, was die Abtastrate anbelangt, für alle Audioanwendungen mehr als ausreichend sein. Dass sie das am Ende in der Praxis nicht ganz ist, weil die im Konvertierungsprozess nötigen, heute verfügbaren Antialiasing- und Rekonstruktionsfilter nicht hinreichend steilflankig (bei einer bestimmten Frequenz scharf abschneidend) sind, um im Bereich zwischen 20.000Hz und 22.050Hz das Signal hinreichend zu dämpfen ohne die Frequenzbereiche darunter zu beschädigen oder sog. Pre-Ringing zu verursachen, habe ich in meinem Artikel über die "Digital Audio Grundlagen" ausführlich erläutert.

III. Frequenzbereiche von Musik

Parallel zur Frage, wie weit das menschliche Gehör Frequenzen wahrzunehmen in der Lage ist, ist auch zu fragen, welche Frequenzbereiche überhaupt sinnvoller Weise übertragen werden sollten. Dabei darf vermutet werden, dass das Ziel der HighEnd-Bemühungen die möglichst realitätsnahe Reproduktion von Klängen aus Musik bzw. Sprache ist.

Unter Klang versteht man in der Akustik ein periodisches aus Sinustönen zusammengesetztes Schallereignis, bei dem die Frequenzen der Teiltöne (Grundton und Obertöne) in einem ganzzahligen Verhältnis zueinander stehen, sich also „harmonisch“ zueinander verhalten. Die Obertonreihe eines Grundtones bestimmt generell nicht nur die Klangcharakteristik, sondern hat aufgrund ihrer Zusammensetzung (etwa bezüglich der Unterschiede in der Stärke der einzelnen Obertöne und wie sie sich im Zeitverlauf des Klanges ändern) auch direkten Einfluss auf die Klangfarbe des wahrgenommenen Tones. Dies ist der Grund, warum derselbe Ton, z.B. a', auf einer Oboe völlig anders klingt als auf z.B. dem Klavier. Nicht-harmonische Schallereignisse werden als "Tongemische" und im Grenzfall als "Geräusche" bezeichnet.

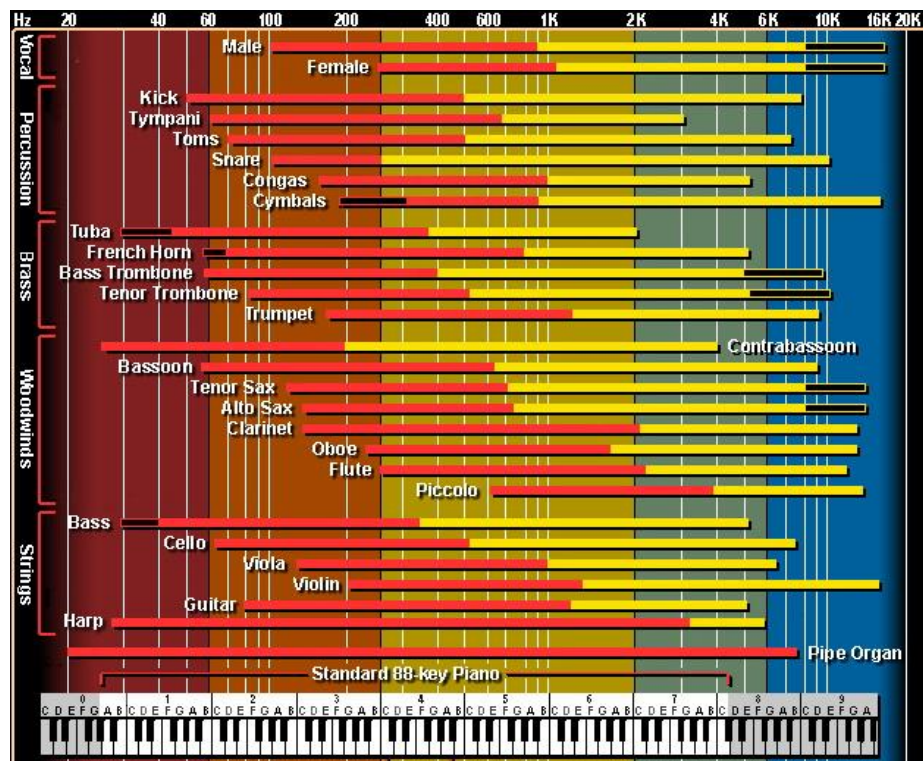
Reale Klänge sind im Allgemeinen komplexer, wobei Geräuschkomponenten und Inharmonizitäten eine wichtige Rolle spielen. Bei der menschlichen Stimme und bei Blasinstrumenten ist die Teiltonzusammensetzung weitgehend harmonisch und befindet sich in guter Übereinstimmung mit dem einfachen Modell. Bei Saiteninstrumenten ist die Frequenzzusammensetzung nur mehr näherungsweise harmonisch, und bei Schlaginstrumenten sind nur noch einzelne Teiltöne fast harmonisch, während das Gesamtspektrum eher unharmonisch ist.

Erst eine weiter gefasste Auffassung von „Klang“ in der Musik nimmt auch die unharmonischen Schallanteile zur Kenntnis, die das musikalische Spiel begleiten, z. B. die Geräusche des Luftwirbels beim Anblasen einer Flöte, das Schnalzen der Saiten auf dem Korpus eines gezupften Streichinstrumentes oder die Schmatz- und Zischlaute der menschlichen Singstimme bei Konsonanten. Somit ist die Klangcharakteristik eines Klages nicht nur vom Instrument, sondern auch vom Musiker und nicht zuletzt auch von der Raumakustik abhängig. All diese Faktoren werden – einzeln oder in ihrer Gesamtheit – allgemeinsprachlich unter dem diffusen Begriff „Klangfarbe“ subsumiert.

Für die Klangfarbe ist natürlich auch das Einschwingverhalten entscheidend, das heißt der zeitliche Verlauf des Spektrums und der Lautstärke in den ersten Sekundenbruchteilen eines Tons. Blendet man diese ersten Zehntelsekunden aus, lassen sich einige Instrumente nur noch schwer identifizieren. Doch interessiert uns dieser Aspekt im Zusammenhang der Abwägung sinnvoller Abtastraten nicht sondern ausschließlich der Frequenzumfang.

All diese Aspekte müssen bei der digitalen Reproduktion realitätsnah erfasst und wiedergegeben werden. D.h., der nutzbare Frequenzumfang muss jedenfalls ausreichend groß sein, um all diese Aspekte völlig fehlerfrei zu erfassen.

In der folgenden Graphik sind die Frequenzbereiche ausgewählter Musikinstrumente und der menschlichen Stimme (näherungsweise) dargestellt, wobei der Grundtonbereich rot und der Obertonbereich gelb dargestellt ist:



Graphik 5: Grund- und Obertonspektren akustischer Musikinstrumente (Quelle: Independent Recording Network)

Auffallend ist, dass es kein akustisches Instrument gibt, dessen Grundtonumfang 8kHz überschreitet. Die Orgel hat zwar das ausgedehnteste Grundtonspektrum, dafür aber kaum Obertonspektren. Selbst die Obertonspektren aller anderen akustischen Instrumente reichen nur in Einzelfällen, wie dem Becken oder der Violine, bis in den Bereich von 16kHz.

Anders liegt der Fall natürlich bei Synthesizern, die in der Lage sind Töne in jeder beliebigen Höhe zu generieren.

Zwischenergebnis: Die letzten zwei Kapitel kann man mit der Aussage zusammenfassen, dass es keine nennenswerten akustischen Schallereignisse oberhalb von 16kHz gibt, die es wert wären, aufgenommen zu werden und dass es kaum Mensch gibt, die in der Lage sind, Töne oberhalb von 18kHz wahrzunehmen, selbst wenn diese extreme Lautstärkepegel aufweisen. Bei 20kHz ist spätestens für alle Menschen absolut Schluss mit der Wahrnehmung auch bei extremsten Lautstärkepegeln.

IV. Sinn und Unsinn hoher Samplingraten

Aufgrund dieser Tatsachen, argumentieren nun viele, vor allem technisch sehr qualifizierte Menschen nicht unplausibel, dass Abtastraten (und auch Wortbreiten) jenseits des CD-Standards unsinnig seien, da sie schlicht unhörbar sind (Stichwort: HiFi für Fledermäuse). Entsprechend hochauflösende Geräte würden das reproduzierbare Frequenzspektrum lediglich in Bereiche erweitern, die für Menschen nicht wahrnehmbar seien und wären damit in etwa so sinnvoll, wie Fernseh-Bildschirme, die infrarotes Licht abzubilden vermögen (Xiph.Org-Stiftung).

Ein vielzitiertes Experiment, das diese theoretischen Überlegungen auch in der Praxis zu unterstützen scheint, ist 2007 von Brad Meyer und David Moran in den USA durchgeführt und im ENGINEERING REPORT der AES Convention vom September 2007 veröffentlicht worden: Mit Hilfe eines ABX CS 5 „double blind“-Comparators wurden die auditive Unterscheidbarkeit zwischen DVD-Audio/SACD kontra 16Bit/44,1kHz-Audio anhand von 60 Testpersonen und 554 Versuchen genauer untersucht. Der Artikel ist leider nur kostenpflichtig zu erwerben: <http://www.aes.org/e-lib/browse.cfm?elib=14195>. Das Gesamtergebnis betrug dabei 49,82% – die Gruppe der Tonschaffenden und Audiophilen erreichte dabei gerade einmal 52,7%. Das Ergebnis belegt, dass keine signifikant sichere Erkennung zwischen dem Audiosignal direkt von SACD/DVD-Audio und demselben Signal, welches über einen 16-Bit/44,1kHz-A/D-D/A-Loop geroutet wurde, möglich war. Der Versuch zeigte gleichzeitig, dass alle auf SACD und DVD-Audio veröffentlichten Stereoaufnahmen in gleicher Audioqualität auch auf einer CD hätten veröffentlicht werden können.

Gleichzeitig haben viele Musikliebhaber über die Jahrzehnte immer wieder die Erfahrung gemacht, dass eine Vinylschallplatte wesentlich besser klingt, als dasselbe Album in CD-Qualität bzw. dass die SACD- oder DVD-A-Version eines Albums deutlich besser klingt, als dasselbe Album auf CD (von Ausnahmen einmal abgesehen, in denen Musiklabels CD-Aufnahmen hochgerechnet und als hochauflösende Tonträger verkauft haben). Im Grunde berichten alle HiFi-Fachzeitschriften weltweit von diesem Phänomen. Zumindest wäre mir keine Fachzeitschrift bekannt, die die Meinung vertritt, dass analoge Vinylaufnahmen und hochauflösende Audiodateien, keinerlei Klangvorteile gegenüber der CD brächten. Nun mag mancher diesen Fachzeitschriften – die immerhin von Menschen betrieben werden, die sehr viel und sehr gut Musik hören, also in bester Position sind, um Klangunterschiede feststellen zu können – unterstellen, dass sie alle nur dem Marketing-Hype der Unterhaltungselektronikindustrie dienen. Aber dererlei Verschwörungstheorien greifen, unsere Ansicht nach, zu kurz. Es ist unplausibel anzunehmen, dass alle Fachzeitschriften weltweit gleichgeschaltet wären. Wenn es

einen großen Bluff gäbe, dann gäbe es auch die eine oder andere Fachpublikation, die dies aus ihrem täglichen Testalltag bestätigen würde. Uns ist aber, wie gesagt, keine solche Fachzeitschrift bekannt.

Was sind also mögliche Erklärungsansätze dafür, dass trotz der technisch hinreichenden Spezifikation des Red-Book-Standards, die meisten Musikliebhaber analoge Tonquellen und hochauflösende Digitaldateien, den CD-Standard-Dateien vorziehen (ungeachtet des Brad-Meyer- & David-Moran-Experimentes)? Ein Großteil der z.T. leidenschaftlich vorgetragenen Argumente basiert sicherlich auf Unkenntnis und Placeboeffekten, denen in neutral durchgeführten ABX-Vergleichstests die Grundlage entzogen werden würde. Darüber hinaus, gibt es aber auch fundierte Gründe, die für höhere Auflösungen sprechen:

- Das Nyquist-Shannon-Theorem setzt ideale Antialias- und Rekonstruktionsfilter voraus, die es in der Realität nicht gibt. Bei einer Abtastrate von 44,1kHz verbleiben nur etwa 2kHz Bandbreite zwischen 20kHz und 22,05kHz für die Anwendung der Filter, die entsprechend steilflankig und dadurch invasiv für das obere Ende des Hörspektrums ausgelegt werden müssen. Sie verursachen, je nach Topologie, Phasenverzerrungen, Pre-Ringing, Höhenabfall oder unharmonische Aliasingverzerrungen im oberen Frequenzbereich. Höhere Abtastraten ermöglichen aufgrund des breiteren verfügbaren Frequenzspektrums den Einsatz deutlich sanfterer Rekonstruktions-Filter, die die Nachteile der Filter bei niedrigeren Sampleraten vermeiden. Dadurch verbessert eine höhere Abtastrate den reproduzierten Klang, nicht weil man das erweiterte Frequenzspektrum hören könnte, sondern weil es eine Anwendung wenig invasiver Antialiasing- und Rekonstruktionsfilter ermöglicht.
- Mit dem Red-Book-Standard kommt neben der Abtastrate von 44,1kHz auch eine Wortbreite von 16Bit, die über den gesamten Dynamikumfang von 96dB lediglich 65.530 Lautstärkeabstufungen ermöglicht. Das menschliche Gehör löst allerdings deutlich feiner auf. Die psychoakustische Forschung legt Auflösungen im Bereich von über 1 Mio. Lautstärkeabstufungen nahe, was einer vergleichbaren digitalen Auflösung von 20-22Bit entsprechen würde.
- Der nicht selten bessere Klang des Musikmaterials auf hochauflösenden Tonträgern, wie DVD-Audio- und SACD gegenüber dem durchschnittlich auf CD veröffentlichten Musikmaterial, liegt nicht zuletzt daran, dass die verantwortlichen Tonmeister und Masteringingenieure für Veröffentlichungen auf SACD und DVD-Audio mehr Freiheiten bekamen, auf hohe Klangqualität zu fokussieren, als dies für CD-Veröffentlichungen üblich ist.

Dem entgegen steht natürlich der zunehmende Speicherbedarf und Processing-Aufwand von Dateien mit höheren Sampleraten. Teilweise wird auch nicht unberechtigt argumentiert, dass hohe Sampleraten ab 176,4Khz und aufwärts sogar schädlich für die Klangqualität der Wiedergabe sind, da sie dem DAC nachgelagerte Gerätschaften, wie Verstärker und Lautsprecher, mit unnötigen hochfrequenten Signalen belasten und so ihre Leistung verschlechtern. Jedoch hängt dies vom konkreten Einzelfall der Auslegung nachgelagerter Gerätschaften ab und ist nicht allgemeingültig.

V. Resümee:

Wie so oft im Leben, gibt es auch bzgl. der Frage nach den optimalen Digitalisierungsparametern (Samplingrate & Wortbreite) keine eindeutigen Antworten. Zuviel hängt von den Spezifika des individuellen Equipments und der verwendeten Filteralgorithmen ab. Dennoch möchten wir denjenigen, die keine Jahre mit Hörvergleichen verbringen wollen, ein paar persönliche Erfahrungswerte an die Hand geben, die unseren Umgang mit Digital Audio prägen:

1. Es kommt wesentlich auf die Qualität der ursprünglichen Aufnahme an (und damit auf die Leistung des Toningenieurs und des Mastering-Ingenieurs). Es gibt hervorragende Aufnahmen in Red-Book-

Standard-Qualität und mittelmäßige hoch-auflösende Aufnahmen und umgekehrt. In Summe sind natürlich zumeist die hochauflösenden Aufnahmen besser, zumindest weil sie sorgfältiger gemastert wurden als dasselbe Album für Veröffentlichung auf CD und weil sie höhere Wortbreiten verwenden.

2. Die Qualität heutiger digitaler Aufnahmen oder Remasterings übersteigen bei weitem die Möglichkeiten heute üblicher - auch high-endiger - Wiedergabeketten (insbesondere Lautsprecher und Raumakustiken), insofern hat der teilweise betriebene Zahlenfetischismus bei Samplingraten zuweilen einen akademischen Charakter.
3. Während wir persönlich bei modernen Aufnahmen keine verlässlichen und nachvollziehbaren Unterschiede bei derselben Aufnahme in unterschiedlichen Sampleraten ab 44,1kHz wahrnehmen können, so dass unserer Meinung nach 44,1kHz oder 48kHz mit heutigen Filtern durchaus ausreichend sind, meinen wir wohl Unterschiede in der Wortbreite derselben Aufnahme ausmachen zu können, wobei unserer Erfahrung nach auch 20Bit-Aufnahmen deutlich besser klingen als 16Bit-Aufnahmen.
4. Etwas anders sieht die Sache bei älteren Aufnahmen (das sind Aufnahmen vor der Jahrtausendwende) aus. Da die früher verwendeten Antialiasing-Filter, invasiver für das Hörspektrum waren, als moderne Filter, bevorzugen wir bei älteren Aufnahmen, möglichst höhere Auflösungen.
5. Hinzu kommt die Frage ob der verwendete Digital-Analog-Converter ein NOS-DAC ohne Digitalfilter ist oder eine Oversampling-Variante mit starken Filtern. Digitale Filter 'verbrauchen' einen Teil der Frequenzbandbreite für ihre Wirkung, so dass man mit filterlosen NOS-DACs auch geringere Samplingraten gut hören kann, während man bei Oversampling-DACs höhere Samplingraten benötigt, damit die Filter ihre Arbeit möglichst wenig invasiv für das Hörspektrum ausüben können.
6. Daraus folgt für uns persönlich, dass wir grundsätzlich versuchen von jeder Aufnahme eine verlustfreie 24Bit-Version zu bekommen. Die Samplerate ist dabei bis 96kHz weitgehend egal. Sind mehrere Versionen, z.B. 48kHz und 96kHz verfügbar, präferieren wir etwas "Headroom" und entscheiden uns für die 96kHz-Version. Darüber hinaus haben wir aufgrund der Dateigrößen ein Präferenz für "nicht zu hohe" Sampleraten, also 88,2kHz/24Bit oder 96kHz/24Bit anstatt 176,4kHz, 192kHz oder gar 384kHz. Wenn es also dieselbe Datei beim Downloaden in 24/96 oder 24/192 gibt, bevorzugen wir die 24/96-Version, die grundsätzlich einen sehr gut Kompromiss aus allen relevanten Parametern darstellt. Diese Einschätzung wird auch u.a. durch die Untersuchungen der DECCA-Records, der EMI/Abbey Road Studios und der Deutschen Grammophon Gesellschaft/Emil Berliner Studios unterstützt, die jeweils ihr gesamtes Archiv an analogen Masterbändern in 24Bit/96kHz digitalisiert haben.
7. DSD-Dateien (2,82MHz/1Bit) halten wir, bei teilweise minimal anderer Klangcharakteristik, qualitativ für vergleichbar mit 20Bit/96kHz-Dateien, wie wir im Artikel "PCM im Vergleich zu DSD" ausführlich dargelegt haben. Zwischen beiden Versionen wären wir indifferent, obwohl wir durchaus subtile Unterschiede in der Klangcharakteristik wahrnehmen, die manchmal zu Gunsten der DSD-Version und manchmal zu Gunsten der PCM-Version ausfallen.

© Alexej C. Ogorek

Quellen:

1. Ellermeier W., Hellbrück J.: Hören - Psychoakustik - Audiologie. In: Weinzierl S. (Hrsg.) Handbuch der Audiotechnik. Springer Verlag, 2008
2. Möser M.: Technische Akustik. 7. Auflage, Springer Verlag, 2007
3. Independent Recording Network