



Diplomarbeit

Untersuchung zur auditiven Differenzierbarkeit digitaler Aufzeichnungsverfahren

Hörvergleich Direct Stream Digital und High-Resolution-PCM (24bit/176,4kHz)

Hochschule für Musik Detmold
Erich-Thienhaus-Institut

vorgelegt von

Dominik Blech

Gartenstr.2
32756 Detmold
Matrikel-Nr.: 6796

und

Min-Chi Yang

Wallgraben 26
32756 Detmold
Matrikel-Nr.: 6980

Studiengang: Musikübertragung (Tonmeisterstudium)

Erstgutachter: Prof. Rainer Maillard

Zweitgutachter: Dipl.-Tonmeister Andreas Meyer

Detmold, den 27. Januar 2004



Fast alle Hard- und Software-Bezeichnungen, die in diesem Buch erwähnt werden, sind gleichzeitig auch eingetragene Warenzeichen und sollten als solche betrachtet werden. Die Verfasser folgen bei den Produktbezeichnungen im Wesentlichen den Schreibweisen der Hersteller.

Alle Rechte vorbehalten. Kein Teil des Werks darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder in einem anderen Verfahren) ohne schriftliche Genehmigung der Verfasser reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Die im Rahmen dieser Diplomarbeit angegebenen Internet-Links waren zum Zeitpunkt der Drucklegung gültig. Für deren Inhalte zu einem späteren Zeitpunkt können die Verfasser aufgrund der ständigen Aktualisierung jedoch keinerlei Garantie übernehmen.

Inhaltsverzeichnis

1	Einleitung	1
2	Technischer Hintergrund	4
2.1	Grundlagen der Analog-Digital-Konvertierung	4
2.2	Der Ansatz von Direct Stream Digital	18
2.3	Vor- und Nachteile von DSD und PCM	21
3	Vorüberlegungen	27
4	Versuchsbeschreibung	30
4.1	Versuchsaufbau	30
4.2	Der Hörraum	36
4.2.1	Versuchsablauf	36
4.2.2	Messungen im Hörraum	39
4.2.2.1	<i>Nachhallzeit</i>	39
4.2.2.2	<i>Hintergrundgeräusch</i>	41
4.2.2.3	<i>Abhörpegel</i>	42
4.3	Informationen zur verwendeten Technik	44
4.4	Versuchsablauf	56
5	Statistische Auswertungsmethode	62
5.1	Die Binomialverteilung	62
5.2	Die Binomialverteilung in Bezug auf den ABX-Hörvergleich	65
6	Auswertung der Versuchsergebnisse	69
6.1	Deskriptive Auswertung	69
6.2	Interpretative Auswertung	76
6.2.1	Allgemeine Vorbemerkungen	76
6.2.2	Wahrnehmungspsychologische Überlegungen	79
7	Fazit	83
8	Danksagung	87
9	Literaturverzeichnis	89
	Erklärung	94

Verzeichnis der Grafiken

Grafik Nr. 1	Kap.2.1	S. 4	Abtastung mit $f_{\text{Abtast}} > 2 \cdot f_{\text{Signal}}$
Grafik Nr. 2	Kap.2.1	S. 5	Abtastung mit $f_{\text{Abtast}} < 2 \cdot f_{\text{Signal}}$
Grafik Nr. 3	Kap.2.1	S. 5	Abtastung mit $f_{\text{Abtast}} = 2 \cdot f_{\text{Signal}}$
Grafik Nr. 4a	Kap.2.1	S. 7	Quantisierungsfehler
Grafik Nr. 4b	Kap.2.1	S. 7	Signal des Quantisierungsfehlers
Grafik Nr. 5	Kap.2.1	S. 9	Quantisierungsfehler bei niedrigem Eingangssignal mit Bildung von Spiegelfrequenzen
Grafik Nr. 6	Kap.2.1	S. 9	Spektrum bei periodischem, diskreten Abtasten im Frequenzbereich
Grafik Nr. 7	Kap.2.1	S.10	Aliasing bei $f_{\text{Abtast}} < 2 \cdot f_{\text{Signal}}$
Grafik Nr. 8	Kap.2.1	S.11	Vermeidung von Aliasing durch Oversampling
Grafik Nr. 9	Kap.2.1	S.12	Digitalsignal ohne korrekten Dither bei Reduktion der Wortbreite
Grafik Nr.10	Kap.2.1	S.13	Digitalsignal mit korrektem Dither bei Reduktion der Wortbreite
Grafik Nr.11	Kap.2.1	S.14	prinzipieller Aufbau eines $\Sigma\Delta$ -Wandlers
Grafik Nr.12	Kap.2.1	S.15	Berechnungsbeispiel für einen $\Sigma\Delta$ -Wandler
Grafik Nr.13	Kap.2.1	S.16	Ausgangssignal eines $\Sigma\Delta$ -Wandlers
Grafik Nr.14	Kap.2.2	S.18	Signalfluss bei Aufzeichnung von PCM
Grafik Nr.15	Kap.2.2	S.18	Signalfluss bei direkter Aufzeichnung eines Ein-Bit-Signals
Grafik Nr.16	Kap.2.2	S.20	Signalfluss bei Aufzeichnung mit mehrbittigem $\Sigma\Delta$ -Wandler
Grafik Nr.17	Kap.2.3	S.22	Impulsantworten bei verschiedenen Samplingraten
Grafik Nr.18	Kap.2.3	S.25	schematische Darstellung des Effekts mehrerer Requantisierungen bei einem DSD-Signal
Grafik Nr.19	Kap.4.1	S.32	Signalfluss-Diagramm des Testaufbaus
Grafik Nr.20	Kap.4.1	S.33	Übersicht der Musik- und Klangbeispiele
Grafik Nr.21	Kap.4.2.1	S.38	schematische Darstellung des Hörraums
Grafik Nr.22	Kap.4.2.2.1	S.40	Toleranzlimit der nominalen Nachhallzeit
Grafik Nr.23	Kap.4.2.2.1	S.40	Nachhallzeiten des Hörraums
Grafik Nr.24	Kap.4.2.2.2	S.41	Hintergrundgeräusch nach EBU/ITU und erreichte Werte im Hörraum - tabellarisch
Grafik Nr.25	Kap.4.2.2.2	S.42	Noise-Rating-Curves NR 10, NR 15 und erreichte Werte im Hörraum - graphisch

Grafik Nr.26a	Kap.4.3	S.44	Frequenzgang Schoeps CMC 6 mit Niere MK4
Grafik Nr.26b	Kap.4.3	S.45	Frequenzgang Schoeps CMC 6xt (MK4)
Grafik Nr.27	Kap.4.3	S.45	Frequenzgang Sennheiser MKH800 (mit Nierencharakteristik)
Grafik Nr.28a	Kap.4.3	S.48	Frequenzgang der MU 2000 der Emil-Berliner-Studios (EBS)
Grafik Nr.28b	Kap.4.3	S.48	Phasengang der MU 2000 der EBS
Grafik Nr.29	Kap.4.3	S.50	Eingangssignal bei einem Sprungantworttest
Grafik Nr.30	Kap.4.3	S.50	theoretisch ideale Impulsantwort der Luft
Grafik Nr.31	Kap.4.3	S.51	Impulsantwort Manger Schallwandler
Grafik Nr.32	Kap.4.3	S.51	Impulsantwort eines Drei-Wege-Systems
Grafik Nr.33	Kap.4.3	S.51	Skizze des Manger Schallwandlers
Grafik Nr.34a	Kap.4.3	S.52	Frequenzgang Zerobox 109
Grafik Nr.34b	Kap.4.3	S.53	Sprungantwort Zerobox 109
Grafik Nr.35a	Kap.4.3	S.53	Frequenz- & Phasengang MSS-Studiomonitor
Grafik Nr.35b	Kap.4.3	S.54	Sprungantwort MSS-Studiomonitor
Grafik Nr.36	Kap.4.3	S.55	Dämpfung des ferromag-Netzfilter-Kabels
Grafik Nr.37	Kap.4.4	S.57	Bildschirmfenster der ABX-Software
Grafik Nr.38	Kap.4.4	S.58	Steuerungsmodul für die ABX-Software
Grafik Nr.39	Kap.4.4	S.59	Fragebogen Nr.1
Grafik Nr.40	Kap.4.4	S.60f	Fragebogen Nr.2
Grafik Nr.41	Kap.5	S.68	Tabelle der Eintrittswahrscheinlichkeiten
Grafik Nr.42	Kap.6.1	S.69	prozentuale Zusammensetzung des Testpublikums nach deren Geschlecht
Grafik Nr.43	Kap.6.1	S.70	Altersstruktur der Testpersonen
Grafik Nr.44	Kap.6.1	S.70	beruflicher Tätigkeitsbereich der Probanden
Grafik Nr.45	Kap.6.1	S.71	Hauptinstrument der Probanden
Grafik Nr.46	Kap.6.1	S.71	Verhältnis von Stereo- und Surroundbeispielen
Grafik Nr.47a	Kap.6.1	S.71	Verteilung von Stereotests auf Musikbeispiele
Grafik Nr.47b	Kap.6.1	S.72	Verteilung von Surroundtests auf Musikbeispiele
Grafik Nr.48a	Kap.6.1	S.72	arithmetisches Mittel der richtigen Antworten für jedes Stereo-Beispiel
Grafik Nr.48b	Kap.6.1	S.73	arithmetisches Mittel der richtigen Antworten für jedes Surround-Beispiel
Grafik Nr.49	Kap.6.1	S.73	Anzahl der Probanden in Abhängigkeit von deren erreichtem prozentualen Ergebnis

1 Einleitung

Seit der Einführung der digitalen Audiotechnik zu Beginn der 1980er Jahre – mit der CD als millionenfach verbreitetem Trägerformat – sind in diesem Bereich enorme technische Fortschritte zu verzeichnen. Damals war es international möglich, sich, beruhend auf Hörversuchen zur Festlegung der nötigen Übertragungsbandbreite, auf einen einheitlichen CD-Standard mit 44,1 kHz bei 16 Bit Wortbreite zu einigen. Dies entsprach in den frühen 80er Jahren den Möglichkeiten von Halbleiter- und Speichertechnologie und galt damals als „state-of-the-art“. Im Zuge der noch immer fortschreitenden Weiterentwicklung – einerseits auf dem Gebiet der A/D- und D/A-Wandlertechnik, sowie im Bereich der digitalen Signalverarbeitung (DSV), andererseits der erfolgreichen Ursachenforschung in Bezug auf digitale Audio-Verzerrungen und darüber hinaus der aus dem Videobereich kommenden Surround-Technik – wurde es erforderlich, einen hochqualitativen Nachfolger für das CD-Format zu entwickeln. Um die enorm angewachsenen Datenmengen aufzunehmen, bot sich die **Digital Versatile Disc (DVD)** mit einer Mindestkapazität von 4,7 GByte als Trägermedium an.

Die (in technischer und klanglicher Diskussion nicht immer einvernehmlich, sondern vielmehr sehr kontrovers verlaufende) Koexistenz von derzeit zwei digitalen Aufzeichnungsverfahren, „**Puls Code Modulation**“ (PCM) und „**Direct Stream Digital**“ (DSD) brachte auch zwei Formate hervor: Die auf PCM basierende **DVD-Audio** und die auf einem 1-Bit-Signal mit 64-facher CD-Abtastrate (=2,8224 MHz) basierende **Super Audio Compact Disc (SACD)**. Sowohl von Hersteller- als auch von Anwender- und Konsumentenseite werden hierbei immer wieder angeblich klangliche und dadurch deutlich wahrnehmbare Unterschiede herausgestellt.

Welches der miteinander konkurrierenden Systeme sich langfristig durchsetzen wird, ist noch immer unklar. Vielfältige technische Untersuchungen und Vergleiche zu diesen Aufzeichnungsformaten werden bereits seit Jahren durchgeführt und im Laufe dieser Zeit des Weiteren unzählige Vorträge bei Tagungen und Seminaren gehalten und „Papers“ veröffentlicht. Einen umfangreichen (offiziellen und den wissenschaftlichen Forderungen nach Reliabilität, Validität und Objektivität genügenden) Hörvergleich unter klanglichen Aspekten, der eigentlich nicht nur bei Tonmeistern oder den

1 Einleitung

sogenannten „Audiophilen“ Interesse wecken sollte, hat es unseres Wissens bedauerlicherweise bislang jedoch noch nicht gegeben.

Wie „klingt“ SACD oder wie „klingt“ DVD-A? „Klingt“ DSD oder PCM – welchen Einfluss hat das digitale Aufzeichnungsverfahren? Was hört man, was will man hören? Fragen, deren Beantwortung eine nähere Auseinandersetzung mit den derzeit aktuellen Diskussionen um die Zukunft von Digital-Audio und dessen Trägern verlangen, denn nicht selten besteht hier auf Konsumenten- und erstaunlicherweise zum Teil auch auf Anwenderseite noch einiger Aufklärungsbedarf.

Gerade aus dem klangästhetischen Blickwinkel der Tonmeister heraus lässt sich bemerken, dass es viele sowohl musikalische als auch technische Faktoren, kombiniert mit dem Wissen um Instrumenten- und Raumakustik, gibt, die den „Klang“ einer Aufnahme ausmachen. Technik lässt sich selbstverständlich bewusst in die Entstehung des Klanges mit einbeziehen und somit stilprägend einsetzen. Ansonsten ist sie (natürlich möglichst perfekt gewünschtes) Mittel zum Zweck – zumindest aus Tonmeister-Sicht. Aber was wird von einem Aufzeichnungsverfahren erwartet? Im allgemeinen eine dem analogen Vorbild möglichst ebenbürtige Darstellung in Bezug auf z.B. Räumlichkeit, Stereo-Abbildung (Surround-Abbildung), Transparenz, Klangbalance, Klangfarbe, Impulsvverhalten, etc. Darüber hinaus jedoch sollte ein digitales Aufzeichnungs- bzw. Kodierungsformat nicht nach etwas „klingen“, sondern vielmehr klangneutral sein. Außerdem muss – auf jeden Fall auf Anwenderseite – ebenso die Frage nach der Praktikabilität im Berufsalltag gestellt werden.

Bei einem ernstzunehmenden Hörvergleich sollte zunächst der Fragestellung nachgegangen werden, was gegenübergestellt werden soll – denn ein solcher Vergleich darf, um später eindeutige Aussagen treffen zu können, nur einen sich unterscheidenden Parameter beinhalten.

Im Rahmen der vorliegenden Tonmeister-Diplomarbeit soll mit Hilfe eines Hörtests eruiert werden, inwieweit Probanden bei einem ABX-Test einen Unterschied zwischen DSD und hochauflösendem PCM (176,4kHz/24Bit) wahrnehmen können. Schon an dieser Stelle sei darauf hingewiesen, dass sich im Laufe des Hörvergleichs immer deutlicher herauskristallisiert hat, wie wichtig ein Testverfahren in Form eines ABX-Tests, eines sogenannten Doppel-Blindtests, ist, da nur so (anhand der Ergebnisse des ABX-Tests) gezeigt werden kann, was noch reproduzierbar wahrnehmbar ist –

1 Einleitung

bekanntermaßen ist der Übergangsbereich zwischen auditiver Perzeption und Imagination sehr schmal.

Im Mittelpunkt unserer Bemühungen standen hauptsächlich zwei Punkte: Einerseits sollte die Signalkette – soweit möglich – auf gemeinsamen Komponenten beruhen, um nicht unterschiedliches Equipment, sondern das Aufzeichnungsverfahren als solches klanglich zu vergleichen, andererseits waren wir bestrebt, einen möglichst breitgefächerten Testpersonenkreis mit unterschiedlichsten Hörerfahrungen, -erwartungen und -fokussierungen zu erhalten.

Nach einem technisch orientierten Kapitel, das die Grundlagen der digitalen Audio-technik und die „Weiterentwicklung“ von DSD, sowie Vor- und Nachteile der konkurrierenden Systeme darlegen soll, und einem Kapitel mit allgemeinen Vorüberlegungen werden der gesamte Versuchsaufbau und die -durchführung ausführlich beschrieben, bevor sich nach der Erläuterung der statistischen Auswertungsmethode die Beschreibung und Interpretation der Ergebnisse und das Fazit anschließen.

Mit der Durchführung dieser Arbeit soll das bestehende Missverhältnis zwischen der umfangreichen theoretischen Datenlage und dem praktischen Nachweis anhand eines schon lange ausstehenden Hörvergleichs verringert werden und hierzu ein objektiver und sachlicher Beitrag geleistet bzw. ein Anstoß für weitere (klangliche) Untersuchungen und Überlegungen gegeben werden.

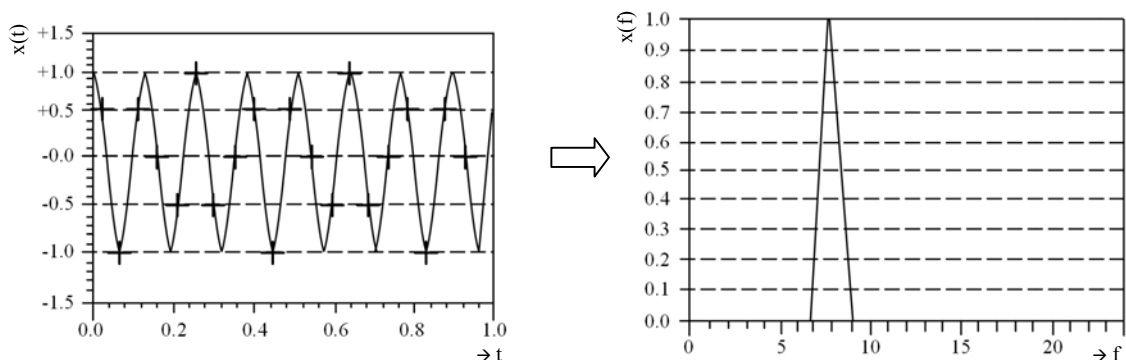
2 Technischer Hintergrund

2.1 Grundlagen der Analog-Digital-Konvertierung

Um ein analoges, d.h. ein zeit- und wertkontinuierliches Spannungssignal zu digitalisieren, muss es in ein zeit- und amplitudendiskretes Signal in Form einer binären Zahlenfolge umgewandelt werden. Die Genauigkeit dieses Konvertierungsprozesses ist von zwei charakteristischen Werten abhängig: Der Abtastrate und der Quantisierung.

Die Abtastrate ist die Frequenz, mit der dem analogen Signal in gleichmäßigen (zeitlichen) Abständen „Samples“ (Proben, Stützstellen) entnommen werden – man spricht deswegen auch von der Samplingfrequenz. Realisiert wird dieses durch die Faltung des Ursprungssignals mit einer periodischen Folge von Dirac-Impulsen, einem zumindest theoretisch unendlich schmalen Rechteckimpuls mit der Amplitude 1. Dadurch entsteht ein zeitdiskretes Signal, welches lediglich zu bestimmten Abtastzeitpunkten definiert ist. Entsprechend des als allgemein bekannt geltenden Shannon'schen Abtasttheorems von 1949 muss der Abtastvorgang mit mehr als der doppelten maximalen Frequenz des Nutzsignales erfolgen, damit das Signal später wieder einwandfrei rekonstruiert werden kann. Es müssen also bei der höchsten abzutastenden Frequenz minimal mehr als zwei Samples pro Periode genommen werden.

$f_{\text{Abtast}} > 2 * f_{\text{Signal}}$: bei einer Abtastung gemäß des Shannon'schen Abtasttheorems lässt sich das ursprüngliche Signal wieder einwandfrei rekonstruieren – siehe Grafik Nr.1.



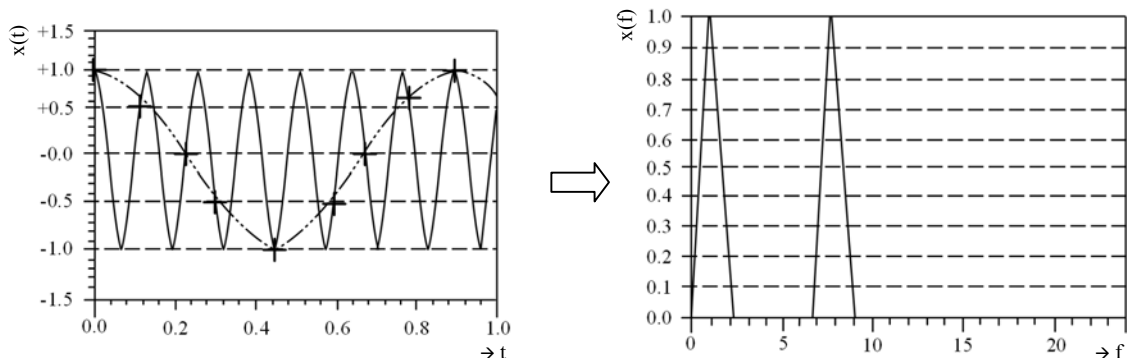
Grafik Nr.1: Abtastung gemäß des Shannon'schen Abtasttheorems.¹

¹ Langen 1997: 7.

2 Technischer Hintergrund

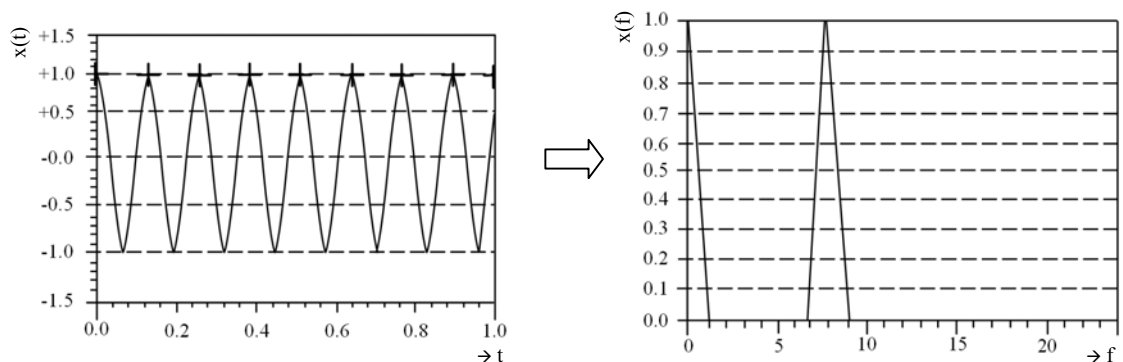
Wird gegen dieses Theorem verstoßen, werden die Samples später falsch interpretiert. Dabei kommen zwei Fälle in Betracht:

$f_{\text{Abtast}} < 2 * f_{\text{Signal}}$: bei der späteren Rekonstruktion des Signals entsteht eine andere (=Alias-) Frequenz als die Frequenz des Eingangssignals – siehe Grafik Nr.2.



Grafik Nr.2: Abtastung mit einer Frequenz kleiner als der doppelten Signalfrequenz: Es entsteht eine andere (Alias-) Frequenz als die ursprüngliche Signalfrequenz.²

$f_{\text{Abtast}} = 2 * f_{\text{Signal}}$: bei der späteren Rekonstruktion des Signals werden die Abtastpunkte falsch interpretiert. Da sie alle identisch sind, werden sie als Stützstellen einer Gleichspannung angesehen – siehe Grafik Nr.3.



Grafik Nr.3: Abtastung mit Signalfrequenz bewirkt DC-Aliasing. Interpretation als Gleichspannung.³

Des Weiteren legt die Abtastrate fest, bis zu welcher Frequenz das zu wandelnde Analogsignal später verarbeitet werden kann, nämlich genau bis zur Hälfte der Abtastfrequenz, der sogenannten Nyquist-Frequenz⁴.

² Langen 1997: 6.

³ Ebd.

⁴ Da das menschliche Hörvermögen im optimalen Falle von 16Hz bis 20kHz reicht, sollte für die Wiedergabe des Audiosignals diese Bandbreite weitestgehend ausgenutzt werden. Die Nyquist-Frequenz sollte somit oberhalb von 20kHz liegen.

2 Technischer Hintergrund

Die im Allgemeinen übliche Samplingfrequenz von 44,1 kHz rührt aus den Anfängen der digitalen Audio-Ära, in denen die „enormen“ Bandbreiten von über 1 MBit pro Sekunde pro Audiokanal gespeichert werden mussten. Um eine brauchbare Aufnahmezeit zu erhalten, griff man auf Videorecorder zurück. Gezwungenermaßen musste die Samplingfrequenz den üblichen Fernsehstandards mit Bild- und Zeilenfrequenzen angepasst werden, so dass ganze Samples pro Zeile gespeichert werden konnten. Bedingt durch den Umstand, dass zwei Fernsehstandards existieren – 525 Zeilen bei 60Hz (M-NTSC) und 625 Zeilen bei 50Hz (BG-PAL) – musste es ein Vielfaches dieser zwei sein, bei gleichzeitiger Eignung als Samplingfrequenz (Abtasttheorem).

Diese Frequenz errechnet sich bei diesem „Pseudo-Videosystem“ aus der Multiplikation von aktiven Zeilen (ohne die ausgetasteten) mit der Bildwiederholfrequenz und der Anzahl der Samples pro Zeile.

Beim NTSC-Verfahren entfallen 35 ausgetastete Zeilen, woraus sich 490 pro Bild, bzw. 245 pro Halbbild ergeben. Bei PAL verringert sich die Anzahl der aktiven Zeilen um 37 auf 588, daraus folgend 294 pro Halbbild. Bei drei Samples pro Zeile ergibt sich:

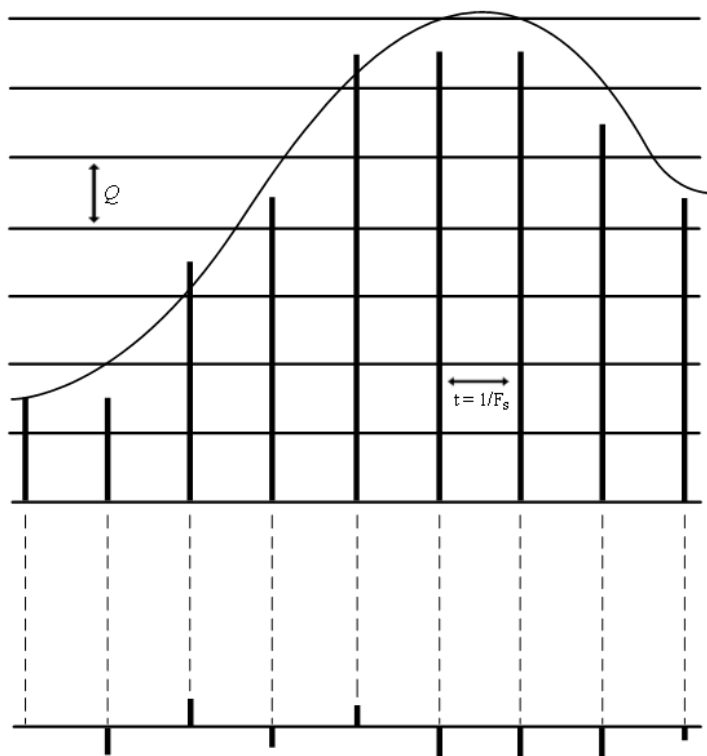
$$\text{NTSC: } 60 * 245 * 3 = 44,1 \text{ kHz}$$

$$\text{PAL: } 50 * 294 * 3 = 44,1 \text{ kHz}$$

Obwohl die CD natürlich keinen Bezug zu Video hat, übernahm man diese Samplingfrequenz, da das Equipment für die Masterherstellung auf Videoequipment basierte.

Bei der Quantisierung entsteht aus dem wertkontinuierlichen ein wertdiskretes (also amplitudendiskretes) Signal. Die Anzahl der Quantisierungsstufen errechnet sich mit 2^n , wobei n die Wortbreite (Anzahl der Bits) ist. Daraus ergibt sich die Genauigkeit der vertikalen Auflösung, mit der die Amplitude der Samples erfasst wird. Die analogen Werte, die zwischen diese Stufenrasterung fallen, werden dem nächstgelegenen diskreten Wert angenähert. Daraus entsteht der sogenannte Quantisierungsfehler, also ein Spannungsfehler, der sich aus der Differenz der Stufenübertragungsfunktion zur analogen Signalkurve errechnet. Dieser Fehler beträgt maximal $\pm \frac{1}{2}Q$ (Q = Quantisierungsstufe) – siehe Grafik Nr.4a:

2 Technischer Hintergrund



Grafik Nr.4a:

Quantisierungsfehler bei einem willkürlichen Signal: Die Abweichung beträgt maximal $\pm \frac{1}{2}Q$. Die Spitzen der Abtastwerte sitzen jeweils in der Mitte des Quantisierungsintervalls.⁵

Grafik Nr.4b:

Die Quantisierungsfehler kann man sich wie ein eigenes, nicht erwünschtes Signal vorstellen, dass dem ursprünglichen hinzugefügt wird.⁶

Arbeitet ein Wandler rauschfrei, so bestimmt die Quantisierung auch den Signal/Rauschabstand (SNR = Signal to Noise Ratio) der Digitalisierung. Im Allgemeinen gilt hier die Faustregel:

$$\text{SNR [dB]} = 20 \log (2^n)$$

Bei einem komplexen und großen Eingangssignal wird sich die Größe des Quantisierungsfehlers mit gleicher Wahrscheinlichkeit zwischen den Grenzen $\pm \frac{1}{2}Q$ verteilen. In diesem Fall entspricht das ungewollte Signal (s. Grafik Nr.4b) einem breitbandigen, spektral weißen (Quantisierungs-)Rauschen, welches mit dem Eingangssignal unkorreliert ist. Es lässt sich folgender SNR berechnen:

Wie oben erwähnt existieren 2^n Quantisierungsintervalle mit einem ebensolchen Spitze-Spitze-Wert der Amplitude (ohne Clipping). Der Spitzeneffektivwert beträgt demzufolge $\frac{2^{n-1}}{\sqrt{2}}Q$. Der Quantisierungsfehler mit einer Amplitude von $\frac{1}{2}Q$ Spitze entspricht einem Effektivwert von $\frac{Q}{\sqrt{12}}$.

⁵ Watkinson 1995: 117.

⁶ Ebd.

2 Technischer Hintergrund

Aus der Formel des Nutz-zu-Störsignal-Verhältnisses SNR lässt sich nun folgern:

$$\text{SNR} = 20 \log_{10} \frac{Y_0}{Y_r}$$

mit Y_0 = Effektivwert des Ausgangssignals ohne
Quantisierungsrauschen

mit Y_r = Effektivwert des Quantisierungsrauschens

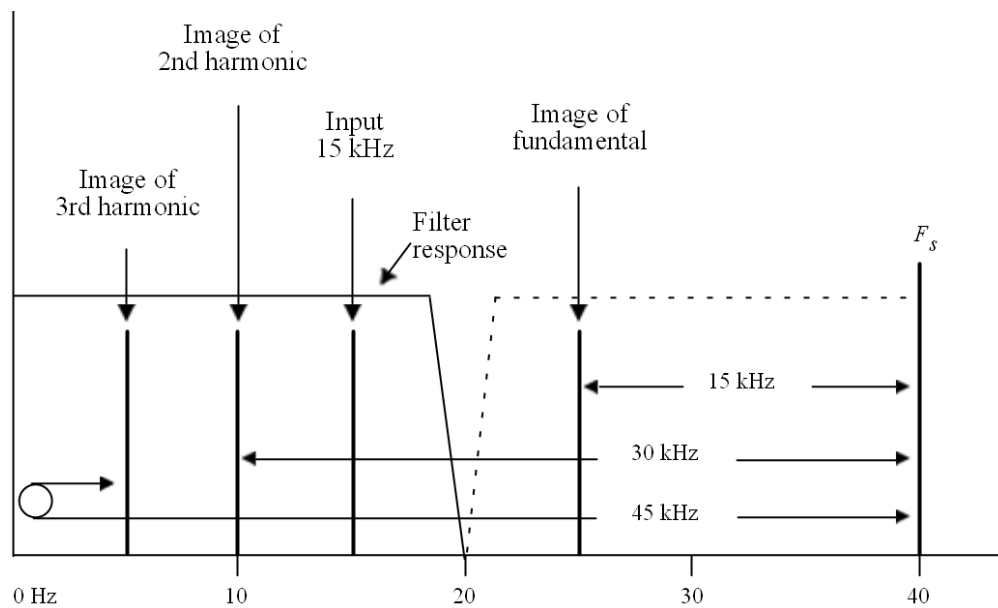
$$\begin{aligned} \text{daraus folgt: } \text{SNR [dB]} &= 20 \log_{10} \frac{\sqrt{12} * 2^{n-1}}{\sqrt{2}} \text{ [dB]} \\ &= 20 \log_{10} (\sqrt{6} * 2^{n-1}) \text{ [dB]} \\ &= 20 \log_{10} \left(2^n * \frac{\sqrt{6}}{2} \right) \text{ [dB]} \\ &= 20n \log_{10} 2 + 20 \log_{10} \frac{\sqrt{6}}{2} \text{ [dB]} \end{aligned}$$

$$\text{SNR [dB]} = 6,02n + 1,76 \text{ [dB]}^7$$

Bei niedrigem Eingangspegel ist die Funktion des Quantisierungsfehlers nicht mehr zufällig, sondern wird von der Wellenform des Eingangssignals abhängig. Das Nutzsignal ähnelt einer Rechteckfolge und der Quantisierungsfehler macht sich als Verzerrung bemerkbar. Diese nichtlinearen Verzerrungen verursachen ein harmonisches Obertonspektrum, welches jedoch erst nach dem Anti-Aliasing-Filter auftritt. Es bilden sich Spiegelfrequenzen (Aliasfrequenzen) und es kommt zu tonalen Artefakten (granulation noise) und Rauschmodulation. – s. Grafik Nr.5:

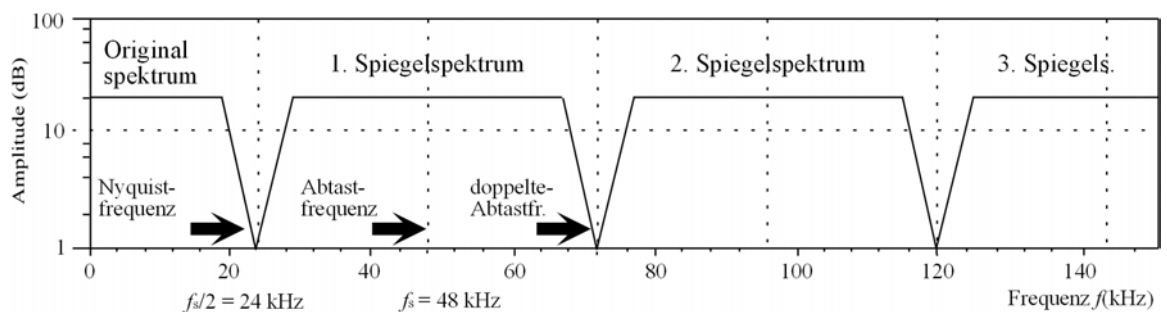
⁷ Watkinson 1995: 118.

2 Technischer Hintergrund



Grafik Nr.5: Bei hohen Pegeln macht sich der Quantisierungsfehler als Rauschen bemerkbar. Je niedriger der Eingangsspiegel jedoch wird, umso korrelierter ist er mit dem Eingangssignal. Das Ergebnis sind Verzerrungen nach dem Anti-Aliasing-Filter. Bei einem Eingangssignal von 15 kHz bilden sich 2. und 3. harmonische Verzerrungen mit Spiegelfrequenzen bei 10 und (-)5 kHz.⁸

Die Anti-Aliasing-Filter (Tiefpass) bewirken, dass die analogen Eingangssignale vor der A/D-Konvertierung auf die maximal zulässige Bandbreite begrenzt werden. Diese ist durch die Nyquist-Frequenz (= halbe Abtastfrequenz) gegeben. Ein periodisches, diskretes Abtasten des Signals im Zeitbereich hat im Frequenzbereich ein ebenso periodisches, diskretes Spektrum zur Folge. Daraus ergibt sich, dass das Originalspektrum sich jeweils mit unterem und oberem Seitenband symmetrisch an der Samplingfrequenz und deren Vielfachen spiegelt – s. Grafik Nr.6:



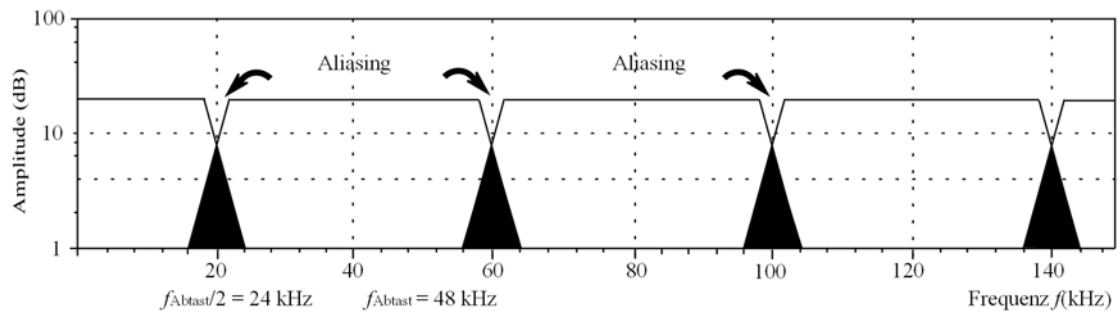
Grafik Nr.6: Abtastung mit doppelter Frequenz des Sperrbereichs des Anti-Aliasing-Filters. Es tritt gerade kein Aliasing auf. Durch periodisches Abtasten des Signals wird das Originalspektrum an der Abtastfrequenz und deren Vielfachen periodisch gespiegelt.⁹

⁸ Watkinson 1995: 119.

⁹ Langen 1997: 8.

2 Technischer Hintergrund

Es wird deutlich, dass – sobald gegen das Abtasttheorem verstoßen wird – die jeweils äußeren Seitenbänder ineinander greifen und es u.a. zu einer Überlagerung des Originalsignals mit dem gespiegelten Spektrum kommt. Um die Nyquist-Frequenz und deren Vielfache entstehen Aliasingfrequenzen. Der in Grafik Nr.7 aufgezeigte Fall tritt auf, wenn die Samplingfrequenz kleiner als die größte doppelte Signalfrequenz ist:



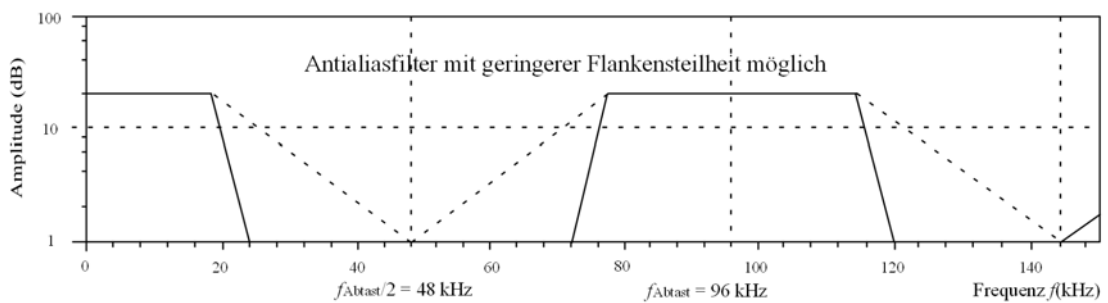
Grafik Nr.7: Abtastung mit einer kleineren Frequenz als der größten doppelten Signalfrequenz.¹⁰

Auf der Grundlage einer Samplingfrequenz von 44,1 kHz muss ein solches Filter also bis 20 kHz möglichst linear übertragen, andererseits aber bis zur Nyquist-Frequenz von 22,05 kHz eine genügend hohe Sperrdämpfung erreichen. Es ergaben sich analoge, extrem steilflankige Filterdesigns achter oder höherer Ordnung mit dementsprechend großen Nachteilen¹¹. Aus diesem Grund erhöhte man die Abtastfrequenz um das 2-, 4- oder 8-fache der Standardabtastrate. Dieser Prozess des sogenannten Oversamplings bewirkt, dass die Spiegelspektren erst erheblich höher einsetzen, somit deutlich weichere analoge Tiefpässe (zweiter oder dritter Ordnung) als Anti-Aliasing-Filter eingesetzt werden konnten. Ein wirksamer Nebeneffekt ist, dass aufgrund der erhöhten Abtastrate der Signal/Rauschabstand (bei gleicher Quantisierung) steigt. Das zuvor erwähnte breitbandige, vom Eingangssignal unkorrelierte Rauschen bei höheren Pegeln verteilt sich auf ein breiteres Spektrum bis zur neuen Nyquist-Frequenz, demzufolge existiert weniger Störleistung im Audioband. Innerhalb der gleichen Bandbreite verringert sich der Störpegel bei jeder Verdoppelung der Samplingfrequenz um 3dB, die spektrale Leistungsdichte halbiert sich – s. Grafik Nr.8:

¹⁰ Ebd.

¹¹ Bei derartig steilflankigen Filtern wird die Phase stark gedreht und die Gruppenlaufzeit ist immer frequenzabhängig. Deswegen benötigen die unterschiedlichen Frequenzen unterschiedliche Zeiten, um die Filter zu passieren (Problem der Dispersion). Daraus resultierend kann der Klang sich eventuell verfärbten. Des Weiteren tendieren analoge Filterdesigns solch hoher Ordnung zu Instabilitäten (Schwingungen) und sind natürlich von Temperatur, Toleranzen der Bauteile sowie deren Alterungserscheinungen abhängig.

2 Technischer Hintergrund



Grafik Nr.8: Abtastung bei wesentlich größerer Frequenz als der höchsten vorkommenden Signalfrequenz, so dass keine Aliasingfrequenzen entstehen.¹²

Des Weiteren ergab sich im Gegensatz zu analogen Tiefpässen bei den verwendeten digitalen FIR-Filtern (*Finite Impulse Response-Filter*) in den Oversamplern der Vorteil der Linearphasigkeit, daraus resultierend konstante, frequenzunabhängige Gruppenlaufzeiten, d.h. alle Frequenzen verzögern sich gleichmäßig in den Filtern. Im Zuge der derzeitigen (v.a. technischen) Diskussionen um DSD bzw. PCM gerieten genau diese linearphasigen Anti-Aliasing-Filter ins Kreuzfeuer der Kritik. Auf diesen Punkt wird in Kapitel 2.3 genauer eingegangen.

Die nichtlinearen Verzerrungen, die bei niedrigem Eingangssignal auftreten, betreffen den Bereich des LSB, des *Least Significant Bit*, die letzte Stelle des digitalen Ausgangswortes – bei der CD also den ungefähren Bereich bei -90dBFS (dB Fullscale). Damit ein Wandler auch hier vollkommen verzerrungsfrei arbeitet, wird dem Eingangssignal vor der Quantisierung ein Dither-Rauschen mit bestimmten statistischen Eigenschaften hinzuaddiert, um die treppenstufenartige Übertragungsfunktion, die bei geringen Pegeln entsteht, zu glätten. Ziel ist, dass der Quantisierungsfehler einen rauschartigen und signalunabhängigen Charakter annimmt.

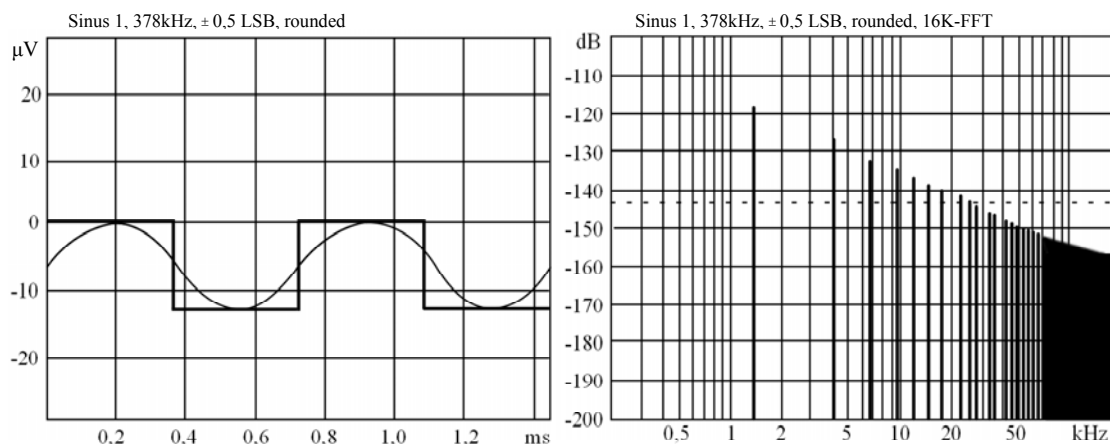
Einwandfreies Dithering ist ebenfalls bei jedem Requantisierungsschritt innerhalb der DSV-Kette, spätestens bei der Wortlängenkürzung (der sogenannten Truncation) der D/A-Konvertierung notwendig: Hierbei dürfen auf keinen Fall die niederwertigeren Bits einfach „abgeschnitten“ werden, weil dies wiederum zu Verzerrungen bei kleinen Pegeln führen würde, sondern auch hier muss der auftretende Quantisierungsfehler durch Hinzufügen von Dither vom Signal dekorreliert und in ein signalunabhängiges weißes Rauschen transformiert werden.

¹² Langen 1997: 9.

2 Technischer Hintergrund

Als Pioniere und bedeutende Vertreter auf diesem Gebiet sind sicherlich Stanley Lipshitz, John Vanderkooy und Robert Wannamaker von der University of Waterloo, Kanada zu nennen, die als erste die Notwendigkeit von Dither mit dreiecksverteilter Wahrscheinlichkeitsdichte (TPDF = Triangular Probability Density Function) zur Unterdrückung von Quantisierungsverzerrungen und Rauschmodulation propagierten. Praktikabel sind für die digitale Audiotechnik allerdings lediglich „nicht-subtraktive“ Ditherverfahren.

Anhand der unten gezeigten Abbildungen wird deutlich, wie notwendig das Addieren von Dither ist: Bei einem 20-Bit-D/A-Konverter wird von einem Sinussignal mit einer Amplitude von $\pm 0,5$ LSB (bezüglich des D/A-Konverters) und 8-fach Oversampling ausgegangen. Durch eine einfache Rundung (von einer zuvor erheblich höheren Wortbreite) entsteht eine reine Rechteckschwingung mit einer Amplitude von 1 LSB; mit der Frequenz des ehemaligen Sinussignals schaltet das niederwertigste Bit alternierend hin und her, dementsprechend entsteht im Frequenzbereich (rechts) das typische Oberwellenspektrum einer Rechteckschwingung. Eingezeichnet ist als gestrichelte Linie außerdem der Pegel des analogen Ausgangsrauschens eines guten 20-Bit-Konverters. Sämtliche Oberwellen, die ins Audioband fallen, liegen oberhalb des Rauschens und sind tatsächlich wahrnehmbar¹³ – s. Grafik Nr.9:



Grafik Nr.9: Digitalsignal ohne korrektes Dithering beim Übergang auf eine geringere Wortbreite.¹⁴

Wird nun vor der Rundung ein geeignetes Dither-Rauschen hinzugefügt (im Fall der Messung in Abbildung Nr.10: $\pm 0,5$ LSB mit gleichverteilter Amplitudendichte¹⁵), lassen

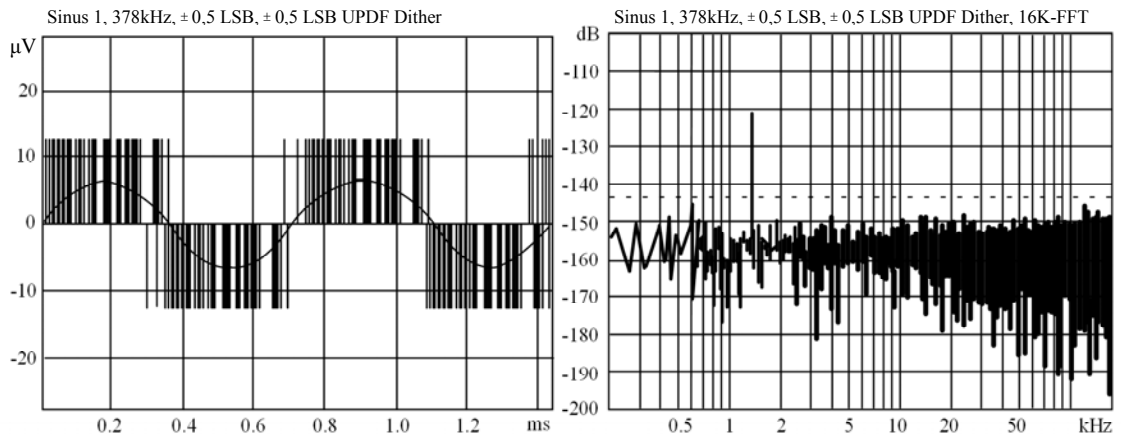
¹³ Müller 2002: 37.

¹⁴ Ebd.

¹⁵ Zur vollständigen Beseitigung aller korrelierten Artefakte wäre streng genommen ± 1 LSB dreiecksverteilter Amplitudendichte erforderlich.

2 Technischer Hintergrund

sich dadurch die Oberwellen eliminieren. Man erhält stattdessen ein weißes Grundrauschen, welches deutlich unter dem analogen Ausgangsrauschen des 20-Bit-Konverters liegt. Bei einem korrekt gedithertem Digitalsystem versinken verstummende Töne ebenso verzerrungsfrei im Rauschteppich wie bei einem analogen System¹⁶ – s. Grafik Nr.10:



Grafik Nr.10: Digitalsignal mit korrektem Dithering beim Übergang auf eine geringere Wortbreite.¹⁷

Die technische Realisierung von Analog-Digital-Wandlern findet in der heutigen digitalen Audiotechnik hauptsächlich durch das Sigma-Delta-Verfahren statt. Die Idee des Oversamplings weiterführend, wird mit 64-facher Standard-Abtastfrequenz gearbeitet ($64 \cdot 44,1 \text{ kHz} = 2,8224 \text{ MHz}$), allerdings nur mit einem Bit. Der resultierende Signal/Rauschabstand eines einfachen Ein-Bit-Wandlers wäre aber so schlecht – obwohl sich die Rauschleistung bis zur Nyquistfrequenz bei 1,4112 MHz verteilen würde – dass man den gewünschten Dynamikumfang nur durch das sogenannte Noise-Shaping-Verfahren erreichen kann. Hierbei wird das statistisch gesehen weiße Rauschen im Spektrum so gelegt, dass davon möglichst wenig auf das Audioband entfällt.

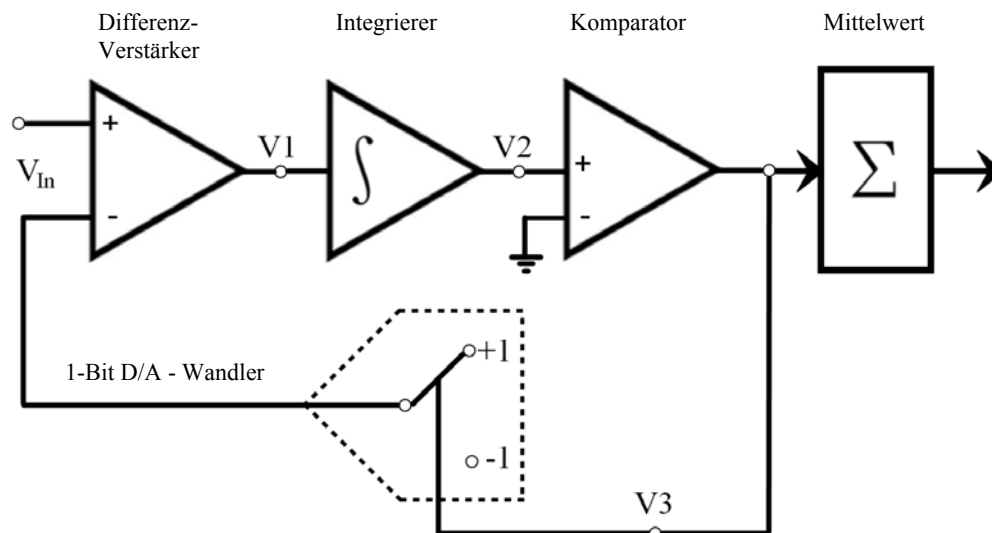
Ein Sigma-Delta-Wandler ($\Sigma\Delta$ -Wandler, s. Grafik Nr.11) besteht grundsätzlich aus zwei Stufen, einem analogen Modulator und einem digitalen Filter. Der analoge Modulator setzt sich prinzipiell zusammen aus einem in Reihe geschalteten Eingangsdifferenzverstärker, einem Integrierer (Σ) und einem Komparator (Δ), der als 1-Bit-Quantisierer fungiert, und einem Ein-Bit-D/A-Wandler im Gegenkopplungsweig. Das

¹⁶ Müller 2002: 38.

2 Technischer Hintergrund

nachgeschaltete Filter (digitales Dezimierfilter) dient der anschließenden Mittelwertbildung aus einer Anzahl von Abtastwerten.

Die grundsätzliche Funktionsweise lässt sich am besten anhand eines Sigma-Delta-Wandlers erster Ordnung zeigen (die Ordnung wird bestimmt durch die Zahl der Tiefpässe oder Integrierer vor dem Komparator): Das Eingangssignal soll dafür $\pm 1V_{SS}$ betragen und an den Punkten V1, V2 und V3 betrachtet werden.



Grafik Nr.11: Prinzipieller Aufbau eines Sigma-Delta-Wandlers.¹⁸

Durch den Eingangsdifferenzverstärker wird vom Eingangssignal das gegengekoppelte Ausgangssignal des 1-Bit-D/A-Wandlers subtrahiert, zu messen am Punkt V1. Dieses wird im Integrierglied zu dessen vorherigem Ausgangssignal hinzuaddiert (Messpunkt V2) und anschließend im Komparator mit einer Spannung von 0V verglichen. Ist V2 größer als 0V nimmt er am Ausgang (=V3) einen Spannungswert von +1V an, sollte er kleiner als 0V sein, legt er den Ausgang auf -1V. Im Gegenkopplungszweig schaltet der getaktete Ein-Bit-D/A-Wandler entsprechend des Messpunktes V3 auf + bzw. -1 Volt und gibt diesen Wert an den Eingangsdifferenzverstärker weiter, wo er wiederum vom nun anliegenden eventuell veränderten Eingangssignal V_{In} abgezogen wird. Dieser Zyklus wiederholt sich pro Abtasttakt.

¹⁷ Ebd.

¹⁸ Langen 1997: 14.

2 Technischer Hintergrund

Die nachfolgende Tabelle in Grafik Nr.12 soll als Berechnungsbeispiel diesen Konvertierungsvorgang mit Hilfe eines Gleichspannungswertes von +0,6V verdeutlichen:

Takt – Nummer	V1	V2	V3	Mittelwert
0	±0,0	±0,0	±0,0	1/8 Σ P3 = 0,5 Volt
1	+0,6	+0,6	+1,0	
2	-0,4	+0,2	+1,0	
3	-0,4	-0,2	-1,0	
4	+1,6	+1,4	+1,0	
5	-0,4	+1,0	+1,0	
6	-0,4	+0,6	+1,0	
7	-0,4	+0,2	+1,0	
8	-0,4	-0,2	-1,0	

Grafik Nr.12: Berechnungsbeispiel des Ausgangssignals eines $\Sigma\Delta$ -Wandlers 1.Ordnung bei einer anliegenden Gleichspannung von $V_{in} = +0,6V$.¹⁹

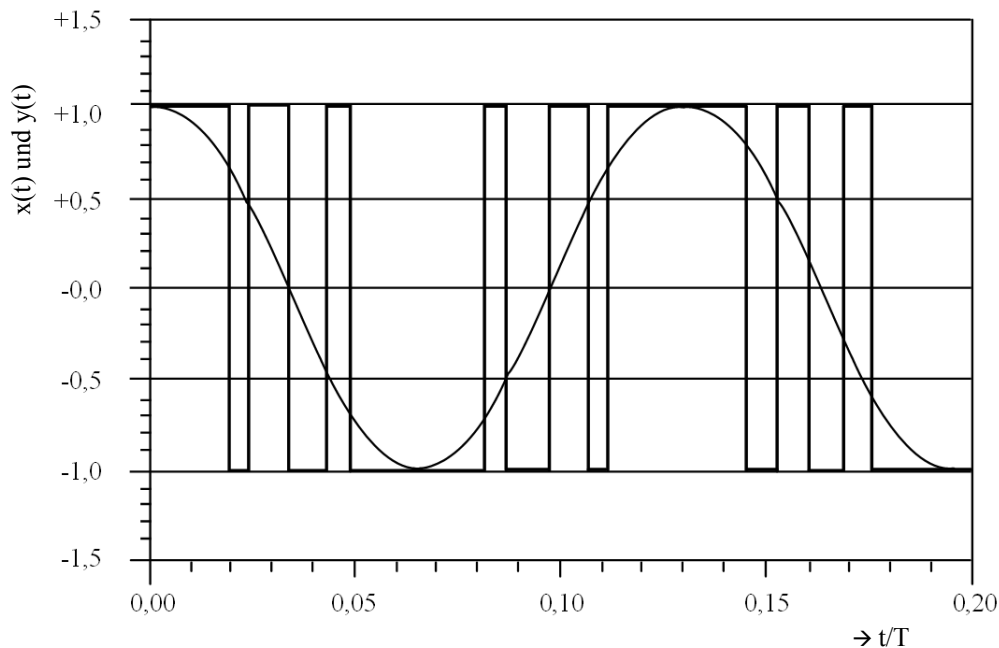
Zunächst liegen alle Messpunkte auf 0V. Gibt man die Gleichspannung von +0,6V nun im ersten Abtasttakt auf den Eingangsdifferenzverstärker, übernimmt er diesen Wert auf den Ausgang an V1, da der D/A-Wandler noch auf 0V liegt. Das Integrierglied addiert diese 0,6V zu seinem am Ausgang liegenden Wert von noch 0V. An V2 ist daraufhin eine Spannung von ebenfalls 0,6V zu messen. Der Komparator legt nun seinen Ausgang V3 auf +1V, da seine Eingangsspannung größer Null ist. Der D/A-Wandler gibt diesen Ausgangswert weiter an den Differenzverstärker, der, beginnend mit dem zweiten Abtasttakt, die +1V von der weiterhin als Eingangssignal anliegenden Gleichspannung von +0,6V abzieht. Es ergeben sich daraus -0,4V an V1, usw.

Mit Hilfe des darauffolgenden, die Abtastrate dezimierenden, digitalen FIR-Filters, das prinzipiell wie ein mittelwertbildender Akkumulator funktioniert, wird das digitale Ausgangswort mit der gewünschten Samplefrequenz und Wortbreite erzeugt.

Das am Ausgang des $\Sigma\Delta$ -Modulators anliegende Signal ist ein 1-Bit-Datenstrom, der einem pulswertenmodulierten Signal ähnelt – s. Grafik Nr.13:

¹⁹ Langen 1997: 16.

2 Technischer Hintergrund



Grafik Nr.13: Ausgangssignal eines $\Sigma\Delta$ -Modulators bei sinusförmigem Eingangssignal.²⁰

Wie schon erwähnt, ist die rein aus dem Prinzip des Sigma-Delta-Wandlers (durch Bitanzahl = 1 und 64-faches Oversampling) sich ergebende Dynamik zu gering, um den Ansprüchen von Audiosignalen zu genügen. Das am Ausgang des Modulators anliegende Signal enthält einen extrem hohen Anteil an Quantisierungsrauschen. Um den Signal-Rauschabstand zu erhöhen, wird dieses durch das Noise-Shaping-Verfahren spektral geformt und zum größten Teil aus dem Audioband in den unhörbaren Bereich geschoben. Die Verformung des Quantisierungsrauschens erfolgt bereits im Integrierglied, somit also schon im analogen Bereich. Dies lässt sich mathematisch bei Darstellung des Sigma-Delta-Wandlers als Linearem Modell herleiten²¹. Wenn man den Integrierer als analogen Tiefpassfilter realisiert, ergibt sich ein Ausgangssignal, das aus der Summe des tiefpassgefilterten Eingangssignals und des hochpassgefilterten Rauschens besteht. Somit erfüllt er einerseits die Anti-Aliasing-Filterung, andererseits aber auch die spektrale Verformung des Quantisierungsrauschens.

Das nachfolgende digitale Dezimierungsfiler verrichtet ebenso wie das Integrierglied zwei Aufgaben: zum einen wird das Quantisierungsrauschen mitsamt aller Audioanteile oberhalb der halben Samplingfrequenz herausgefiltert, so dass eine Unterabtastung auf die gewünschte Samplingfrequenz möglich ist (Abtasttheorem), zum anderen dient das

²⁰ Langen 1997: 15.

²¹ Langen 1997: 18.

2 Technischer Hintergrund

FIR-Filter zur Erzeugung der erforderlichen Wortbreite des Ausgangssignals. Dies geschieht durch das Zusammenfassen mehrerer verzögerter 1-Bit-Samples zu einem n -Bit-Sample. Die dadurch erzielbare Wortbreite ist nicht von der Dynamik des Sigma-Delta-Modulators abhängig, sondern von der Filterfunktion.

Technisch realisiert wird dieses digitale Filter häufig durch ein Brickwall-Filter, da der Frequenzgang dieses Filters nahezu einer angenäherten Rechteckfunktion entspricht. Wie auch oben schon erläutert, ist der Vorteil dieses FIR-Filters (*Finite Impulse Response-Filter*) v.a. der lineare Phasengang und damit eine frequenzunabhängige Gruppenlaufzeit. Entworfen als nichtrekursives Filter (ohne Rückkopplungspfad), ist es weniger empfindlich gegenüber Quantisierungseffekten als ein IIR-Filter (*Infinite Impulse Response-Filter*).

Auf der D/A-Seite setzte man zu Beginn der digitalen Audio-Ära überwiegend Wandler nach dem R-2R-Prinzip ein. Dieses Konverterprinzip, auf das hier nicht weiter eingegangen werden soll, zeichnet sich zwar auch heutzutage noch durch ein sehr geringes Grundrauschen und demzufolge einen sehr hohen Dynamikumfang aus, konnte sich jedoch aufgrund v.a. sehr hoher Bauteilkosten²² nicht durchsetzen. Zum Einsatz kommt wiederum der Sigma-Delta-Wandler. Hierfür wird das Audiosignal zuerst auf üblicherweise das 8-fache der Standard-Abtastrate überabgetastet (Oversampling mit Hilfe eines Interpolationsfilters), nach dem $\Sigma\Delta$ -Wandler steht anschließend wieder das Ein-Bit-Stream-Signal zur Verfügung. Von einem internen Switched-Capacitor-Filter wird es abschließend analog vorgefiltert.²³

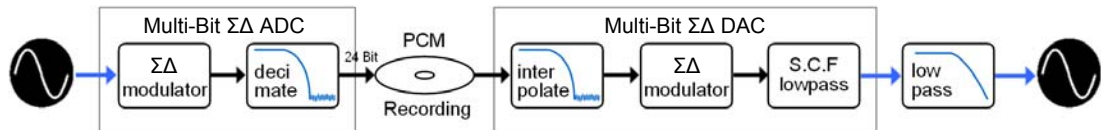
²² Hier sind v.a. möglichst gleiche, deswegen lasergetrimmte Widerstände, daraus resultierend differentielle Nichtlinearitäten und „individuelles“ Klirrverhalten der einzelnen R-2R-Wandler zu nennen.

²³ Müller 2001: 23-24.

2 Technischer Hintergrund

2.2 Der Ansatz von Direct Stream Digital

Man vergegenwärtige sich anhand der unten gezeigten Grafik Nr.14 noch einmal den Signalfluss bei der Aufzeichnung von PCM:



Grafik Nr.14: Signalfluss bei der Aufzeichnung von PCM.²⁴

Zunächst findet unter Verwendung eines Sigma-Delta-Wandlers die A/D-Konvertierung mit hoher Abtastrate statt. Anschließend kann durch Dezimierung mit einem digitalen Filter die gewünschte Quantisierung bei entsprechender Samplingfrequenz erreicht werden. Nach dem Aufzeichnungsvorgang ist für die D/A-Rückkonvertierung zunächst eine Interpolation (Oversampling) mit recht steilflankigen Filtern und erneute Sigma-Delta-Wandlung erforderlich, bevor das Signal als analoges Ausgangssignal zur Verfügung steht.

Hier setzt nun die Idee von Direct Stream Digital (DSD) an: Ließe man das Dezimationsfilter, das Oversamplingfilter und den Sigma-Delta-Modulator einfach weg, und würde das Ein-Bit-Datensignal geradewegs aufzeichnen sowie direkt zu den analogen Tiefpässen auf der D/A-Wandler-Seite schicken, könnte man nicht nur einige Bauteile einsparen, sondern v.a. die steilflankigen Anti-Aliasing-Filter vermeiden. Auch wenn die Speicherung des ungefilterten DSD-Datenstroms mehr Kanalkapazitäten als notwendig erfordert, bedeutet der Wegfall natürlich eine sichtliche Vereinfachung und stellt einen naheliegenden Ansatz dar:



Grafik Nr.15: Signalfluss bei direkter Aufzeichnung des Ein-Bit-Signals.²⁵

²⁴ Nach Müller 2001: 25.

²⁵ Ebd.

2 Technischer Hintergrund

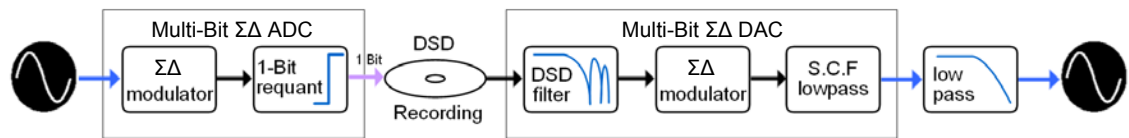
Durch die technische Weiterentwicklung der A/D- und D/A-Konverter als Multi-Bit-Wandler lässt sich dieser zunächst nachvollziehbare Ansatz jedoch mittlerweile in Frage stellen. Den Vorteilen der preiswerten Fertigung und einer sehr guten Linearität von Sigma-Delta-Wandlern steht der deutlich geringere Dynamikumfang sowohl auf A/D- als auch auf D/A-Seite (im Gegensatz zu den erwähnten R-2R-Konvertern) gegenüber. Verantwortlich hierfür ist Jitter – minimale Schwankungen des Abtasttaktes – die sofort zu einem Anstieg des Rauschpegels führen. Hier offenbart sich der Gegensatz: Setzt man sehr viele Bits ein, so steigen die Verzerrungen, bei wenigen Bits sinken diese. Bezüglich Jitter verhält es sich genau invers. Bei hoher Bitzahl ist Jitter ein erheblich geringeres Problem als bei wenigen Bits. Der Wunsch nach einer Minimierung dieses Gegensatzes und nach einer Erhöhung des Signal-Rauschabstandes konnte nur durch variierte Wandler erfüllt werden: es kamen schnelle 4- oder 5-Bit-Konverter statt der einfachen Komparatoren zum Einsatz, wodurch das notwendige Noise Shaping weniger stark ausgelegt werden konnte. Um differentielle Nichtlinearitäten möglichst zu vermeiden, werden in den IC's mehrere dieser Wandler eingesetzt, die das Signal jeweils per Zufall erhalten. Der Fehler, der sich aus der Nichtlinearität eines einzigen Wandlers ergibt, wird geglättet und macht sich als Rauschen bemerkbar.

Der kritische Leser wird bemerken, dass das bei DSD erforderliche Ein-Bit-Signal im Signalfluss nun gar nicht mehr auftaucht, da als natives Format nach dem Sigma-Delta-Modulator jetzt ein Multi-Bit-Signal zur Verfügung steht. Damit aus dem (bereits digitalisierten) Ausgangssignal des Modulators das Ein-Bit-Signal gewonnen werden kann, ist ein nachgeschalteter digitaler Ein-Bit-Requantisierer erforderlich. Diese notwendige Stufe beinhaltet natürlich wiederum das Hinzufügen von zusätzlichen Verzerrungen und zusätzlichem Rauschen. D/A-seitig ist zwar die Berechnung des 4- bis 5-bittigen Modulator-Eingangswortes durch Tiefpassfilterung nötig, dies ist jedoch verzerrungsfrei möglich. Die digitale Vorfilterung ist sogar recht günstig, da hier bereits hochfrequente Störleistung beseitigt wird.²⁶

²⁶ Müller 2001: 24ff.

2 Technischer Hintergrund

Insgesamt gesehen ergibt sich damit folgende Grafik Nr.16:



Grafik Nr.16: Signalfluss bei der heutzutage üblichen Aufzeichnung mit Hilfe mehrbittiger Sigma-Delta-Wandler.²⁷

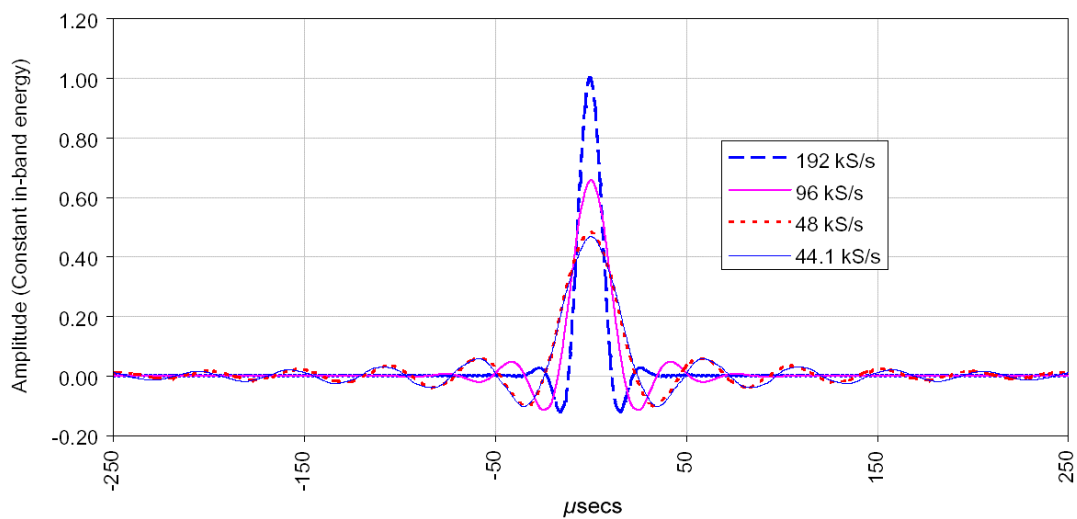
²⁷ Nach Müller 2001: 25.

2.3 Vor- und Nachteile von DSD und PCM

Die vor allem technisch orientierten Diskussionen in Bezug auf jeweilige Vor- bzw. Nachteile von DSD und hochauflösendem PCM dauern bereits seit mehreren Jahren an und sind mit diversen Fachaufsätzen, Convention Papers, Artikeln und Interviews belegt. Im folgenden Kapitel soll versucht werden, einen kleinen und neutralen Überblick über die jeweils positiven und negativen Aspekte der unterschiedlichen Aufzeichnungssysteme zu geben. Erklärtes Ziel beider Aufzeichnungsverfahren ist selbstverständlich die Steigerung der digitalen Audioqualität.

Hauptkritikpunkt an PCM sind für die DSD-Befürworter die Anti-Aliasing-Filter. Diese verursachen den sogenannten Filter-Ringing-Effekt: Linearphasige Filter haben eine symmetrische Impulsantwort, sie weisen also nicht nur ein Aus- sondern auch ein spiegelsymmetrisch zur Ankunftszeit des Signals liegendes, ebenso langes Einschwingen auf. Dabei sind beide Schwingungsvorgänge umso länger, je höher die Ordnung, also je steilflankiger das Filter ist (s. Grafik Nr.17). Bereits vor dem Eintreffen eines transienten Impulses reagiert das Filter mit einem Einschwingvorgang. Genau diese Signalverfälschungen sollen nach Aussage der DSD-Befürworter – auch noch bei Abtastraten bis 192 kHz – zu hören sein und bei sehr kurzen Transienten mit einem sehr ausgebreiteten Spektrum zu „Verschmierungen“ führen. Von der „PCM-Seite“ wird diese Aussage (und der sich aus dieser ergebende, vermeintlich *hörbare* Vorteil von DSD) aufgrund der deutlich erhöhten Abtastraten der DVD-A angezweifelt und mit den – um den Faktor der Abtastratenerhöhung – erheblich verringerten Filterlängen und dem damit verbundenen Zeitraum des Vor- und Nachschwingens argumentiert. Ein großer Vorteil der höheren Abtastrate (96 bis 192 kHz) liegt für sie neben der größeren Audiobandbreite vor allem in der Flexibilität, die nun bei der Entwicklung der benötigten digitalen Filter besteht, sowie der Erleichterung bei der digitalen Signalverarbeitung. Ein weiterer Kritikpunkt der DSD-Befürworter sind auf der A/D-Seite die Dezimations- (Downsampling) und auf der D/A-Seite die Interpolationsfilter (Oversampling), da diese Requantisierungsrauschen hinzufügen.

2 Technischer Hintergrund



Grafik Nr.17: Symmetrische Impulsantworten bei verschiedenen Samplingraten.²⁸

Je höher die Samplingfrequenz, umso niedriger sind die Artefakte durch die Zeitquantisierung. Deswegen propagierten Sony und Philips mit DSD den einbittigen, ungefilterten Datenstrom, der direkt nach dem Komparator des Sigma-Delta-Modulators abgegriffen werden kann. Bei 64-facher Standardabtastrate, also 2,8224 MHz Samplingfrequenz (2,8224 Megasamples pro Kanal pro Sekunde), gilt im Audioband mit Hilfe von Noise Shaping ein Mindestdynamikumumfang von 120 dB und eine Bandbreite bis 100 kHz als SACD-Standard.

Aufgrund der hohen Nyquistfrequenz stellt DSD sehr genügsame Anforderungen an die Anti-Aliasing-Filter. Deren Sperrdämpfung muss erst bei ungefähr 1,4 MHz erreicht werden. Bedingt durch den weichen Filterverlauf oberhalb von 95 kHz ist das Filter-Ringing im Zeitbereich erheblich geringer gegenüber 96 oder auch noch 192 kHz, DSD zeigt hier die geringsten Artefakte.

Der hierfür zu zahlende Preis ist das enorme Quantisierungsrauschen, welches durch die extreme Überabtastung mit 2,8224 MHz Samplingfrequenz entsteht. Da es jedoch nur hochfrequente Komponenten enthält, die mit dem Audiosignal unkorreliert sind, ist es nicht wahrnehmbar.²⁹

Von der Gegenseite wird dahingegen herausgestellt, dass eben dieser neben dem eigentlichen Audiosignal existierende „Quantisierungsschmutz“ wirkungsvoller Unterdrückungsmaßnahmen bedarf. Das starke Noise Shaping verursacht schon ab 20 kHz einen äußerst steilen Anstieg des Rauschpegels, der bei ca. 70 kHz ein extrem hohes

²⁸ Story 1997: 4.

²⁹ Reefman/Janssen 2002: 11.

2 Technischer Hintergrund

Niveau erreicht – insgesamt beinhaltet das DSD-Signal deutlich mehr Energie (acht mal so viel) als ein Sinussignal, welches mit Vollaussteuerung übertragen wird.³⁰ Inwieweit bei diesem enormen Störpegel Subtilitäten und schwachpegelige Details innerhalb des Oberwellenspektrums und von Transienten bei Musikinstrumenten noch zuverlässig wahrgenommen werden können, bleibt auf der Basis des aktuellen Forschungsstandes unklar.

Die Notwendigkeit der Erhöhung der Übertragungsbandbreite begründet sich u.a. durch das breite Spektrum von sehr kurzen Transienten. Bei Entfernen von deren höherfrequenten Anteilen würde sich eine Verbreiterung in eine gauss-förmige Kurve ergeben. Um dieses zu vermeiden, benötigt man eine Bandbreite von 50 bis 100 kHz; für das Unterdrücken der oben erwähnten Ringing-Effekte ganz allgemein sehr breite Übertragungsbänder. Beide Phänomene zusammengenommen, ist ein Übertragungsbereich bis 50 kHz und ein sich daran anschließender Übergangsbereich von 100 kHz – also ein Frequenzbereich bis 150 kHz – nötig, um sämtliche eventuell hörbaren Artefakte zu beseitigen.³¹

Von der Verschwendung an Kanalkapazität bei der Speicherung des nicht unterdrückten Quantisierungsrauschens und der daraus folgenden Unterlegenheit gegenüber des vergleichbaren Multibit-PCM-Modus der DVD-A³² einmal abgesehen, ist bei DSD wiedergabeseitig in den nachfolgenden Analogstufen wegen der Gefahr von Intermodulationsprodukten unterhalb von 20 kHz, Instabilitäten der Endverstärker oder der Zerstörung der Hochtöner eine effektive Tiefpassfilterung absolut notwendig. Neben den digitalen Filtern in den D/A-Konvertern kommen deswegen zusätzlich moderate, analoge Tiefpässe in den SACD-Playern mit Eckfrequenzen zwischen 40 bis 80 kHz zum Einsatz, wodurch die behauptete Bandbreite von 100 kHz und die hohe Impulstreue natürlich zu einem gewissen Grad egalisiert werden. Die hohe Abtastrate von 2,8224 MHz bei DSD liefert somit in der Praxis letztendlich nur ein Passband, dessen linearer Frequenzgang nicht über 40-50 kHz hinausgeht. Zusätzlich nimmt das Signal-Rauschverhältnis oberhalb von 15 kHz bereits deutlich ab.

Bereits in mehreren und allseits bekannten Beiträgen³³ auf AES-Conventions haben Stanley Lipshitz und John Vanderkooy auf deutliche Nachteile der 1-Bit-Sigma-Delta-Modulation hingewiesen. Sie vertreten vehement den Einsatz von Multi-Bit-Modula-

³⁰ Müller 2001: 33.

³¹ D.Reefman in Müller 2001: 45.

³² Gerade bei Gebrauch von effektivem Noise Shaping (Müller 2001: 33).

³³ Hier sind u.a. die AES-Convention Preprints 5188, 5395 und 5398 zu nennen, s. Literaturverzeichnis.

2 Technischer Hintergrund

toren, da sich deren einbittige Brüder bei korrektem Dither permanent im Clipping befinden würden. Dies liegt in der Tatsache begründet, dass allein die Amplitude des TPDF-Dithers (Triangular Probability Density Function, s. Kapitel 2.1), der für die vollständige Linearisierung des Quantisierers zuständig ist, Clipping bewirkt. Wird dann noch das eigentliche Eingangssignal und das Signal der Fehler-Rückkopplung (Error-Feedback) des Noise-Shaping-Quantisierers addiert, trägt dies unweigerlich zu einem weiteren Pegelanstieg bei. Die Konsequenz wären Rauschmodulation, Instabilitäten und Verzerrungen. „Selbst wenn der 1-Bit-Konverter so weit gedithert wird, wie es eben möglich ist, können sich Grenzyklen-Artefakte bis zur Nyquist-Frequenz herauf ausbilden [...]“.³⁴

Des Weiteren ergibt sich ein Problem bei der Bearbeitung von 1-Bit-Signalen, da bei jedem noch so trivialen Signalverarbeitungsschritt (wie z.B. einer Pegeländerung) das 1-Bit-DSD-Datensignal in ein Multi-Bit-Wort konvertiert und anschließend wieder requantisiert werden muss. Sollte nun eine Sequenz von DSV-Schritten notwendig sein, würde der Signalpfad sehr viele Requantisierungen enthalten. Hierzu äußert sich Derk Reefman:³⁵

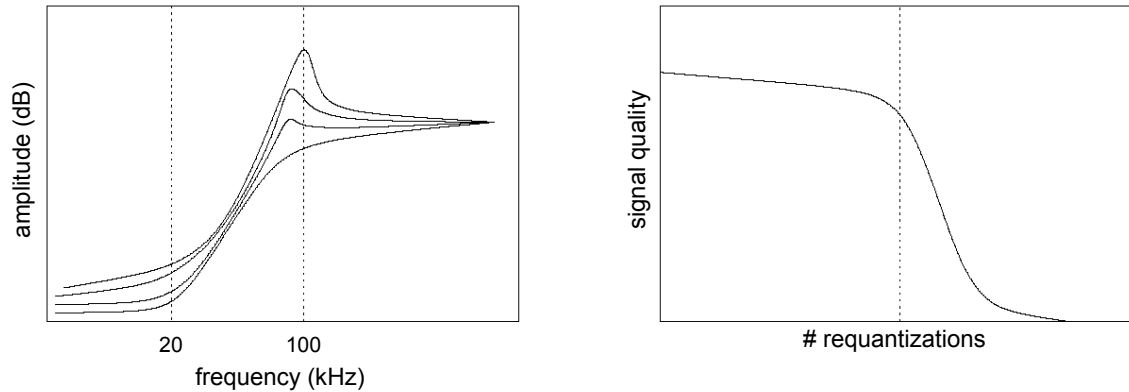
„[...]the total signal path will contain multiple requantizations. As a result of this, build-up of HF-noise will occur. This effect is illustrated in Fig.21 (=Grafik Nr.18, linkes Bild), where schematically the effect of multiple requantizations is displayed. This figure can be explained as follows. If we have a DSD signal, its noise starts to rise above 20-30 kHz, and reaches an almost flat level at about 90 kHz. If, in a subsequent requantization, the bandwidth of DSD is maintained, the signal is low pass-filtered at a frequency of about the same value (90 kHz). If this signal is fed to a next SDM (Sigma-Delta-Modulator), its output signal will contain both its own quantization noise, as well as the quantization noise that has been input to it. If this cascade is repeated, it is easy to see why there will be a build-up of HF-noise in the area of about 80-90 kHz. Eventually, this signal will be large enough to drive the SDM into its clippers, or, worse: instability. This effect is shown in the right of Fig.21 (=Grafik Nr.18); as the number of requantizations increases, the signal quality drops slowly. At the moment that the HF noise is large enough to activate the clippers, the signal quality drops rapidly.

³⁴ St. Lipshitz in Müller 2002: 41.

³⁵ Reefman, Janssen 2002: 33.

2 Technischer Hintergrund

Hence, *all* signal processing should be done in a multi-bit domain; only after the final signal processing step the conversion to $64f_s$ 1-bit signals should be made.”



Grafik Nr.18: Schematische Darstellung des Effektes mehrerer Requantisierungsschritte bei einem DSD-Signal.³⁶

Bereits im Vorwort desselben Aufsatzes stellen die beiden Autoren deutlich heraus.³⁷

“[...] In this respect, it is essential to realize that DSD at $64f_s$ is a *consumerformat* – hence, not necessarily the format that is used in the studio which can be in principle any format as long as it is of equal or better quality compared to standard DSD.”

Da ein Multi-Bit-Signal prinzipiell ohne Verluste bearbeitet werden kann, empfiehlt Philips bis zur endgültigen 1-Bit DSD-Konvertierung die Verwendung von 32-Bit-PCM mit 352,8 kHz Abtastrate bei der gesamten Verarbeitung und Speicherung für die professionelle SACD-Produktion. Auch Sony rät zum Multi-Bit-Ansatz, wobei ihrer Meinung nach jedoch das sogenannte DSD-Wide verwendet werden soll, welches auf einer 8-Bit Version von DSD basiert.

Von den DVD-A-Befürwortern wird die SACD mit ihrer Bandbreite und Dynamik oftmals mit einer DVD-A von 20 Bit und 96 kHz Samplingfrequenz verglichen, welche jedoch den geringen Noise-Floor über die gesamte Bandbreite bis 48 kHz beibehalten kann. Gerne wird darüber hinaus betont, dass auch bei der DVD-A aufnahmeseitig Noise Shaping verwendet werden und somit bei einer Quantisierung von nur 16 Bit der Rauschpegel der SACD (bei einem deutlich geringeren Gesamttrauschpegel) mühelos unterboten werden kann.

³⁶ Ebd.

³⁷ Reefman, Janssen 2002: 7.

2 Technischer Hintergrund

Abschließend sei natürlich der effektive, mehrfache Kopierschutz der SACD genannt, der durch die sukzessiv steigenden Zahlen an Raubkopien in Mitleidenschaft gezogenen, gesamten Audiobranche eventuell zur Regeneration verhelfen könnte. Zum Einsatz kommt hierbei eine Kombination von physikalischen und kryptographischen Methoden. Das oft benutzte Argument der Hybrid-SACD trägt selbstverständlich nicht, da diese Sandwich-Herstellung natürlich bei der DVD-A – wie ja auf dem Markt bereits gezeigt wurde – ebenso möglich ist.

Bei allem Gesagten bleibt noch zu erwähnen, dass bezogen auf die Rezeptivität schneller Transienten, einer erweiterten Audiobandbreite, „Verschmierungen“ bei Filtern, etc. – obwohl auf diesem Gebiet auch schon Versuche durchgeführt wurden – noch immer viele Fragen offen bleiben bzw. auditive Nachweise noch immer ausstehen.

3 Vorüberlegungen

Im Rahmen der vorliegenden Diplomarbeit soll herausgefunden werden, inwieweit ein Unterschied zwischen hochauflösenden, digitalen Aufzeichnungsverfahren respektive DSD und High-Resolution-PCM bei einer Samplingfrequenz von 176,4 kHz³⁸ und einer Quantisierung von 24 Bit noch wahrgenommen werden kann. Bei einem solchen Hörversuch müssen einige grundlegende Aspekte berücksichtigt und dementsprechend Vorüberlegungen angestellt werden – bezogen auf die Auswahl des Equipments, den Hörraum, die zuvor aufgenommenen und dargebotenen Musik- und Klangbeispiele und ein breitgefächertes Testpublikum. Diese Überlegungen sollen im Folgenden dargestellt werden.

Da der Hörvergleich auf das hinter dem reinen Tonträgerformat oder dessen Medium stehende digitale Kodierungsverfahren zurückgeführt werden soll, ist es notwendig, Aufnahme- und Wiedergabepfade zu realisieren, die – soweit möglich – auf gemeinsamen Audio-Komponenten basieren, um nicht unterschiedliches Equipment, sondern das Aufzeichnungsverfahren als solches klanglich zu vergleichen. Der unvermeidbare Schwachpunkt befindet sich bei den A/D- und D/A-Konvertern – eine Minimierung kann daher ausschließlich durch die Verwendung eines Wandlers desselben Herstellers und desselben Modells geschehen, welches beide Kodierungsverfahren unterstützt. Ein Hörvergleich mit Wandlern unterschiedlicher Hersteller hingegen ist mit der Intention der vorliegenden Diplomarbeit nicht sinnvoll zu vereinbaren, da jedes Gerät, bedingt durch analoge Ein- und Ausgangsstufen, eine eigene Klangcharakteristik besitzt – man erhalte Testresultate mit nur schwer einzugrenzenden, unterschiedlichen Parametern, deren Auswertung zu keiner eindeutigen Aussage führen würde.

Weiterhin muss ein praktikabler Testablauf gewährleistet sein, der unnötige Latenzen beim Umschalten zwischen den beiden Ursprungsquellen (DSD und PCM) vermeidet. Dieser ist nicht gegeben, wenn man zwei unabhängige Audio-Workstation-Systeme miteinander verkoppelt, da die Synchronisation immer eine geringe, jedoch deutlich

³⁸ An dieser Stelle sei angemerkt, dass sich die PCM-Samplingfrequenz von 176,4kHz aufgrund der Notwendigkeit der Synchronisation über Word Clock von PCM- und DSD-Wandlern ergab: Die Samplingfrequenz der beiden Kodierungsverfahren müssen einem gemeinsamen Vielfachen von 44,1kHz entsprechen (PCM = 4 mal 44,1kHz; DSD = 64 mal 44,1 kHz).

3 Vorüberlegungen

wahrnehmbare Verzögerung – bei z.B. einer Play- oder Skip-Funktion – für den im Slave-Betrieb arbeitenden Rechner bedeuten würde. Das Differenzieren zwischen den Quellen auf rein zeitlicher Ebene wäre denkbar und möglich. Dementsprechend muss eine Lösung gefunden werden, die das Speichern der beiden unterschiedlichen Aufzeichnungsformate auf nur einer Audio-Workstation gewährleistet. Hier bietet sich das sogenannte AES3-Data-Mapping-Verfahren an. Solch ein Verfahren „verpackt“ bei der A/D-Konvertierung den DSD-Datenstrom in einem Non-Audio-Format vollkommen verlustfrei in 24Bit/44,1kHz-Daten³⁹. D/A-seitig werden die Daten nach demselben Algorithmus wieder als DSD-Signal „entpackt“.

Aufgrund der beiden genannten Forderungen – nach einem Wandler, der ebenso beide Kodierungsverfahren unterstützt wie einen praktikablen Testablauf garantiert – fiel die Entscheidung zugunsten der äußerst hochwertigen Konverter der Firma dCS Ltd., da bei ihnen die Optimierung beider Wandlungsverfahren im Vordergrund steht und sie dem Anwender der A/D- und D/A-Konverter neben dem erwähnten P3D-Modus außerdem die Möglichkeit bieten, zwischen mehreren Filtern zu wählen.

Um einerseits der durchschnittlich größeren Hörerfahrung des potentiellen Testpublikums mit Stereowiedergabe gerecht zu werden, andererseits aber auch eventuell wahrnehmbare Auswirkungen auf räumliche Komponenten mit einzubeziehen, sollte der Hörvergleich sowohl in Stereo als auch in Surround stattfinden. Essentiell hierbei ist, dass die eigens dafür aufgenommenen Musik- und Klangbeispiele absolut identisch sind, folglich in keiner Weise bearbeitet wurden (durch Pegeländerung, Schnitt, Mix, etc.). Aus diesem Grund muss ebenfalls ein klangbeeinflussendes Mischpult umgangen, resultierend daraus das Audiosignal der entweder zwei oder fünf Mikrofone unverfälscht aufgezeichnet und anschließend entsprechend „geroutet“ über die Lautsprecher wiedergegeben werden.

Da jeder musikausübende und/oder -hörende Mensch seine ganz eigenen Hörerfahrungen, -erwartungen und -fokussierungen mit individuellen, instrumentenbezogenen Vorlieben besitzt, sollte nicht nur der Testpersonenkreis, sondern auch das Angebot an zur Verfügung stehenden Musik- und Klangbeispielen möglichst breitbandig gewählt werden. Aus diesem Individualitätsgedanken heraus ist es

³⁹ Bei dCS „P3D-Modus“ genannt: Die DSD-(Stereo)-Daten werden über 3 AES/EBU-Kanäle ausgegeben.

3 Vorüberlegungen

unumgänglich, jeden Probanden einzeln zu testen. Zwar hat das einen erheblich höheren zeitlichen Aufwand zur Folge, aufgrund des sehr eingegrenzten Referenz-Abhörpunktes („Sweet Spot“) und der „ablenkenden Faktoren“, die durch weitere Personen im Hörraum entstehen würden, ist diese Lösung jedoch zu bevorzugen. Es muss durch Schaffung einer möglichst angenehm neutralen räumlichen, wie auch persönlichen Atmosphäre (mit optional einzulegenden Pausen) – ohne vorherige Beeinflussung des Kandidaten – versucht werden, dem Leistungsdruck, dem sich der Proband bei einem subjektiven Test automatisch aussetzt, weitestgehend entgegenzuwirken.

Gesamt gesehen erhebt diese Diplomarbeit den Anspruch, einen aus wissenschaftlicher Sicht einwandfreien Hörvergleich zu präsentieren, der unter möglichst optimalen und den internationalen Richtlinien für einen Hörtest entsprechenden Bedingungen durchgeführt wurde. Die Relevanz des Testverfahrens in Form eines ABX-Hörvergleichs ist dabei nicht zu unterschätzen: Nur so kann – frei von Suggestion und unterbewusster Voreingenommenheit – gezeigt werden, was wir noch wahrzunehmen in der Lage sind. Das Testverfahren des ABX-Tests und die Auswertung der Ergebnisse auf der Grundlage eines nachvollziehbaren Stochastikmodells, im vorliegenden Fall der Binomialverteilung, ist unseres Erachtens nach die geeignetste Möglichkeit, die nicht ganz zu vermeidende „Subjektivität“ zu eliminieren.

Es versteht sich von selbst, dass der gesamte Testaufbau unter Beachtung absoluter Pegelgleichheit zwischen beiden Signalzweigen durchgeführt werden muss.

4 Versuchsbeschreibung

4.1 Versuchsaufbau

Wie schon im vorausgegangenen Kapitel erwähnt, ist eine grundsätzliche Bedingung für einen objektiven und technisch einwandfreien Hörvergleich, dass das zu vergleichende Musikmaterial vollkommen identisch und „unbehandelt“ ist. Das heißt, es darf weder im Pegel geändert, noch darf es editiert, verhallt oder auf eine andere Weise bearbeitet werden (in diesem Fall würde DSD rechnerintern in ein PCM-Signal gewandelt werden). Da solches Musikmaterial – wenn überhaupt vorhanden – nicht zugänglich war, wurde dieses zusammen mit Instrumentalisten der Hochschule für Musik Detmold als Stereo- und Surroundbeispiele in der dem Erich-Thienhaus-Institut angegliederten „Neuen Aula“ eigens vor Beginn der Hörtest produziert.

Um ein klangbeeinflussendes Mischpult umgehen zu können, wurden die Stereo-Musikbeispiele mit je zwei, Surroundbeispiele mit je fünf Mikrofonen (mit erweitertem Frequenzgang bis 40 bzw. 50 kHz) der Firmen Schoeps (MK2S, MK4 und MK41 mit den Mikrofonverstärkern CMC 6-Ug xt) und Sennheiser (MKH800) aufgenommen – folglich war also jedem Lautsprecher ein Mikrofon zugeordnet. Die mit Lake People-Mikrofonvorverstärkern des Typs MIC AMP F/35II auf Line-Pegel angehobenen Signale wurden über 50 Meter lange, niederkapazitive Mikrofonkabel der Firma Klotz (Typ der Serie M1) in den Regieraum geführt, dort mittels eines Y-Adapters aufgesplittet und über dCS-Konverter (insgesamt 2 mal 3 zweikanalige dCS 904) einerseits in DSD, andererseits in PCM (24bit,176,4kHz) A/D-gewandelt und nach dem Datenbitmapping-Verfahren der Konverter mit einem Pyramix „Virtual-Studio-System“ (Merging Technologies) gespeichert. Dies ermöglichte, unabhängig von „Live-Musik“ und dem sich daraus ergebenden Problem der nur „ungenauen Reproduktionsfähigkeit“, Musik vollkommen identisch wiederzugeben und so für den Test nutzen zu können. Für das Bitmapping wurden aufzeichnungsseitig pro DSD-Kanal 3, pro PCM(176,4kHz/24Bit)-Kanal 4 Kanäle mit 44,1kHz/24Bit-Format benötigt. Daraus ergab sich für das Pyramix-System eine Gesamtpurenzahl von 38 Kanälen (die „freie“ Spur des DSD-A/D-Center-Konverters musste „leer“ mitlaufen). Anschließend wurde das Audiosignal ebenfalls über dCS-D/A-Wandler (insgesamt 2 mal 2 zweikanalige dCS 954 für L, R, LS, RS und 2 mal 1 zweikanaligen dCS 955 für C) in ein analoges Signal rückkonvertiert und auf eine von den Emil-Berliner-Studios entwickelte, hochwertige

4 Versuchsbeschreibung

Stereo- und Surround-Monitorunit vom Typ MU 2000 gegeben. Auf diese konnte der Proband mittels einer ebenfalls bei den Emil-Berliner-Studios entwickelten ABX-Software zugreifen und jeweils zwischen den Signalen DSD und PCM auswählen. Mit Hilfe eines softwareseitig im ABX-Test implementierten, auf den Umschaltmoment zwischen zwei Signalen bezogenen Delays, wurde ein gleichzeitiges Erklingen beider Signale und somit eine den Probanden irreführende Pegel-Kaskadierung unterbunden. Als Wiedergabesystem wurden Schallsysteme der Firma Manger eingesetzt, die sich durch eine hervorragende Impulsgenauigkeit und einen Frequenzgang bis 35kHz auszeichnen. Hatte der Proband sich für ein Stereobeispiel entschieden, konnte er darüber hinaus über einen Kopfhörer der Firma Stax hören. Die gesamte Verkabelung wurde ausnahmslos mit neuen, hochwertigen Analog- und Digitalkabeln der Firma Klotz ausgeführt.

Folgende von dCS für diesen Hörvergleich empfohlenen Filtereinstellungen wurden gewählt:

- A/D-DSD-Konverter: Die DSD-Filter unterscheiden sich in der Stärke, mit der sie das Quantisierungsrauschen im Audioband bis 20kHz unterdrücken und wie sie das Quantisierungsrauschen oberhalb des Audiobandes formen. Es wurde Filter 5 empfohlen. Zwar unterdrücken Filter 6 und 7 noch effektiver das Quantisierungsrauschen, jedoch würde der SNR innerhalb des Audiobandes dann nicht mehr im Bereich der SACD-Spezifikationen liegen.
- A/D-PCM-Konverter: Diese Filter sind Anti-Aliasing-Filter und beeinflussen den Frequenzbereich oberhalb von 20kHz. Filter 1 ergibt dabei den steilsten „Cut off“; mit aufsteigender Filternummer werden die Anti-Aliasing-Filter im Übergangsbereich immer „flacher“, wodurch die Impulsantwort jedoch immer präziser und das symmetrische Vor- und Nachschwingen geringer wird. Empfohlen wurde Filter 2.
- D/A-DSD-Konverter: Die unterschiedlichen Filter beeinflussen in Form einer wählbaren Abschwächung der HF-Anteile die Bandbreite des Ausgangssignals und bieten somit eine Anpassung an die Möglichkeiten der Monitoring-Kette. Hier sollte die niedrigste Filternummer, bei der keine Artefakte wahrnehmbar sind,

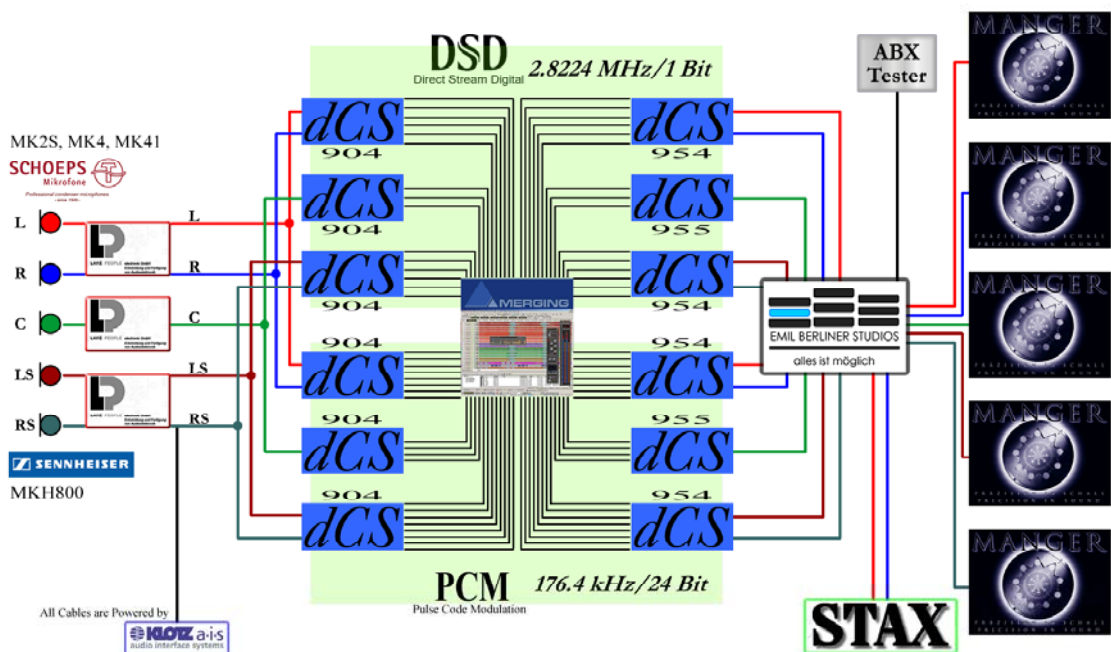
4 Versuchsbeschreibung

genommen werden. Dementsprechend empfohlen und auch gewählt wurde Filter 1.

D/A-PCM-Konverter: Von diesen linearphasigen FIR-Filtern, sogenannten „Anti-Image-Filtern“, stehen bei allen Samplingfrequenzen vier verschiedene zur Auswahl. Sie beeinflussen das Ultraschall-Spektrum oberhalb des Audiobandes und vermeiden Nyquist-Imaging. Die Filter haben nach Filternummer aufsteigend breitere Übergangsbreiten. Empfohlen wurde Filter 1.

Die Auswahl des gesamten Equipments erfolgte ausschließlich nach technischen Gesichtspunkten – nähere Informationen können dem Kapitel 5.3 entnommen werden, der Signalfluss in der Übersicht der folgenden Grafik:

Mikrofon - LakePeople Mikrofonvorverstärker - dCS A/D Konverter 904 - Merging Pyramix Virtual Studio - dCS D/A Konverter 955/954 - EBS Monitor Unit & ABX Tester - Manger Schallsysteme



Alle eingetragenen Warenzeichen sind Eigentum der jeweiligen Firmen und Organisationen.

Grafik Nr.19: Signalfluss-Diagramm des Testaufbaus.

Um den unterschiedlichen Hörerfahrungen, -erwartungen und -fokussierungen auf musikalischer Seite gerecht werden zu können, sollte dem möglichst breitgefächerten Testpublikum während der Durchführung der Hörversuche selbstverständlich ein entsprechend vielfältiges Angebot an Musik- und Klangbeispielen zur Verfügung stehen. Alle parallel in DSD und in PCM (176,4kHz/24Bit) durchgeführten Aufnahmen fanden in der „Neuen Aula“ der Hochschule für Musik in Detmold bei deaktivierter

4 Versuchsbeschreibung

Klimaanlage statt, als Regie diente der Raum 698 („Archiv“). Die folgende Tabelle zeigt eine Übersicht der aufgenommenen Musikbeispiele:

STEREO	SURROUND
Cembalo (Gero Wycik) Fr.Couperin – <i>Rondeau c-Moll</i> 3:21	Cembalo (Gero Wycik) Fr.Couperin – <i>Rondeau c-Moll</i> 3:25
Gesang (Alissa Rose) W.A.Mozart – <i>Le nozze di Figaro, Arie der Susanna</i> „Deh vieni, non tardar“ 3:11	Gesang (Alissa Rose) J.Strauss – <i>Die Fledermaus, Lied der Adele</i> , „Mein Herr Marquis“ 1:32
Gitarre (Tim Büchsenschütz) E.Clapton – <i>Signe</i> 2:06	Gitarre (Tim Büchsenschütz) Unbekannt – <i>Romance</i> 2:20
	Jazz Trio (Christian Struck, Piano – Christoph Terbuyken, Bass – Gabriel Grote, Schlagzeug) M.Manieri – <i>Sarah’s Touch</i> 4:32
Oboe (Pavel Sokolov) G.Ph.Telemann – <i>Phantasie Nr.8, TWV 40:9, e-Moll, 2. Satz (Spirituoso)</i> 1:11	Oboe (Pavel Sokolov) G.Ph.Telemann – <i>Phantasie Nr.8, TWV 40:9, e-Moll, 1. Satz (Largo)</i> 2:22 <i>2.Satz (Spirituoso)</i> 1:14
	Orgel (Hisako Yoshida) M.Reger – <i>Fuge d-Moll, op.135b</i> 5:10
Percussion Solo (Wen-Chi Song) <i>Maracas</i> 1:30 <i>Guiro</i> 1:30 <i>Windchime</i> 1:30 <i>Kastagnetten</i> 1:30	Percussion Trio (Jennifer Fang, Wen-Chi Song, Pi Tzu Yang) N.J.Zivkovic – <i>aus Danza Barbara Tutti-Bsp.Nr.1</i> 2:33 <i>Tutti-Bsp.Nr.2</i> 1:33
Piano (Kimiko Imani) D.Scarlatti – <i>Sonata, K.188 a-Moll</i> 2:38	Piano (Kimiko Imani) Fr.Chopin – <i>Études, op.25, Nr.11 a-Moll</i> 4:04
Sprache – russisch (Ememkut Zaotschnyj) A.Puschkin – <i>aus Eugen Onegin</i> 2:08	Sprache – russisch (Ememkut Zaotschnyj) A.Puschkin – <i>aus Eugen Onegin</i> 2:09
	Streicherorchester (Detmolder Kammerorchester, Ltg. Eckhard Fischer) E.Rautavaara – <i>Pelimannit „Fiddlers“ 2.Satz „Kopsin Jonas“ (Presto)</i> 1:08 <i>5.Satz „Hypyt“ (Presto)</i> 1:16
Trompete (Udo Potratz) <i>Blues Improvisation</i> 2:31	Trompete (Udo Potratz) <i>Blues Improvisation</i> 2:36
	Violine (Jimmy Hsueh) J.S.Bach – <i>Sonate Nr.1, BWV 1001, Adagio</i> 4:19

Grafik Nr.20: Übersicht der zur Verfügung stehenden Musik- und Klangbeispiele.

4 Versuchsbeschreibung

Die Entstehung der Aufnahmen soll an dieser Stelle nicht ausführlich für jedes Beispiel dokumentiert werden, da sie im Rahmen der Arbeit nicht im Mittelpunkt des Interesses steht. Es sollen hier lediglich einige generelle Anmerkungen folgen:

Die Stereobeispiele wurden jeweils klanglich optimiert in AB-Aufstellung entweder mit Schoeps-Mikrofonen (MK2S) bzw. mit Sennheiser-Mikrofonen (MKH800) mit Kugelcharakteristik aufgenommen. Die Verwendung von Mikrofonen unterschiedlicher Hersteller liegt darin begründet, dass die Aufnahmen teilweise parallel sowohl mit Schoeps- als auch mit Sennheiser-Mikrofonen bzw. mit 4 Sennheiser-Mikrofonen gleicher Charakteristik, aber unterschiedlicher Basisbreite durchgeführt wurden (dann nur in DSD oder nur in PCM), um sogenannte „falsche Beispiele“ zu erhalten, mit Hilfe derer während der Tests – im Gegensatz zur Unterscheidungsfähigkeit zweier Aufzeichnungsverfahren – die Differenzierbarkeit von „variiertes Tonmeisterarbeit“ (Auswahl bzw. Aufstellung der Mikrofone) geprüft werden sollte. Von dieser „Irreführung“ wurde während der Testdurchführung jedoch abgesehen, da sich gleich zu Beginn der Tests herausstellte, dass in der Regel ein Proband aus Zeit- und Konzentrationsgründen nur ein einziges, und nicht – wie zuvor geplant – mehrere Beispiele mit jeweils einer kompletten ABX-Testreihe durchlaufen konnte. Dem Probanden wäre es nicht zuzumuten gewesen, den Zeitaufwand eines Testdurchlaufs und der Anfahrt lediglich für ein „falsches Beispiel“ aufzubringen.

Bei den Surround-Beispielen wurde der „Mix“ zwischen L&R, C und LS&RS mit den Lake People-Mikrofonvorverstärkern eingestellt. Im allgemeinen wurde mit 2 Kugeln für L und R, einer Hyperniere als C und 2 in den Raum gerichteten Nieren (jeweils positions-optimiert) mikrofoniert; einzige Ausnahme bilden die Aufnahme des Jazz- und des Percussion-Trios – bei beiden wurden die Mikrofone klanglich optimiert direkt im Ensemble aufgestellt und auf die Spuren/Lautsprecher verteilt.

Die Sprachaufnahme wurde bewusst in russischer Sprache gewählt, da somit eine Maximierung der Konzentration allein auf die Sprachwiedergabe unter Umgehung der Ablenkung durch semantische Inhalte gewährleistet werden konnte. Die Musik- und Klangbeispiele wurden den Probanden in voller Spiellänge zur Verfügung gestellt. Die gesamte Bedienung, wie zum Beispiel Lautstärkeregelung, Zugriff auf die ABX-Testsoftware etc., erfolgte vom Probanden mittels einer Steuerungseinheit, so dass dieser selber über den Ablauf und die zeitliche Einteilung des Hörvergleichs bestimmen

4 Versuchsbeschreibung

konnte. Diese Möglichkeit war ein wichtiger Faktor, um den schon erwähnten Leistungsdruck des Probanden möglichst zu minimieren. Eine genauere Beschreibung des Testablaufs und des ABX-Tests erfolgt im Kapitel 4.4 dieser Arbeit.

Der gesamte Regieraum wurde über das geregelte Netz des Erich-Thienhaus-Instituts und zusätzlich über eine ferromag-Netzfilter-Steckdosenleiste (passive, breitbandige Hochfrequenz-Filterung) versorgt. Eine reibungslose Kommunikation zwischen Tester und Proband war durch eine bidirektionale Sprechverbindung gewährleistet.

Damit sämtliche A/D- und D/A-Konverter unter identischen Bedingungen arbeiten, wurden sie vor Beginn der Tests exakt eingemessen:

Zunächst wurde mit Hilfe des Pegelgebers Tektronix SG 5010 ein 1kHz-Ton mit +6dBu (1,550V) generiert und auf die analogen Eingänge der A/D-Wandler gegeben. Anschließend wurde der Pegel des Digitalausgangs mittels des präzisen Pegelmessers in Pyramix gemessen und auf einen digitalen Headroom von -6dB/Fullscale mit einer Genauigkeit von ca. $\pm 0,01$ dB abgeglichen.

Nachdem die A/D- und D/A-Konverter aneinander angeschlossen waren, wurde mit dem Pegelmesser Tektronix AA 5001 das Ausgangssignal des A/D-Wandlers wiederum auf +6dBu (1,550V) eingemessen, so dass sich in der A/D-D/A-Signalkette ein absoluter 1:1 Durchgang ergab – bei einer Abweichung von ± 2 mV ergibt sich hierbei zwischen analogem Ein- und Ausgangssignal ebenfalls eine maximale Genauigkeit von ca. $\pm 0,01$ dB).

Die A/D-Konverter für DSD können problemlos auf Fullscale PCM eingemessen werden, da die dCS-Wandler diesen Pegel bei Wahl eines DSD-Formats intern automatisch auf den entsprechenden Wert des SACD-Standards gemäß der Definition von 0dB Fullscale übertragen.

Wie im nachfolgenden Kapitel genauer nachgelesen werden kann, wurde als Abhör-raum der Raum 698 im Erich-Thienhaus-Institut eingerichtet und den Richtlinien der EBU/ITU entsprechend akustisch angepasst.

4.2 Der Hörraum

4.2.1 Beschreibung des Hörraums

Als Abhörraum für den Hörvergleich wurde der Raum 698 (das sogenannte „Archiv“) im Erich-Thienhaus-Institut eingerichtet. Um die gering nachteiligen, geometrischen Gegebenheiten zu kompensieren, war es notwendig, einige akustische Anpassungen zur Erfüllung der Richtlinien der European Broadcasting Union (EBU)⁴⁰ und der International Telecommunication Union (ITU)⁴¹ bezogen auf die Nachhallzeit und die Hintergrundgeräusche für Abhörräume vorzunehmen.

Der Raum hat mit seinen Seitenabmessungen von 5,58m x 5,63m x 2,75m (L x B x H) eine nahezu quadratische Grundfläche von 31,4m² und ein Volumen von 86,4m³. Er wird auf zwei rechtwinklig zueinander liegenden Außenwänden von doppelverglasten Doppel-Fenstern (mit einem ca. 15cm breiten Luftzwischenraum) ab ca. 90cm oberhalb vom Fußbodenniveau, und zwei gegenüberliegenden Wänden begrenzt. An den Wänden stehen – jeweils über die gesamte Breite – offene Regalsysteme, die gleichzeitig als Diffusoren und Reflektoren zum Einsatz kamen; die Decke ist holzverschalt.

Zur klanglichen Optimierung wurde die gesamte Fensterfront des Raumes mit Akustikelementen bestückt, wobei abwechselnd einerseits deren stoffbezogene Frontseite als Breitband-Absorber und andererseits deren Rückseite aus massivem Holz als Reflektor verwendet wurde. Der oberhalb der Akustikelemente verbleibende, lichtdurchlassende Spalt von ca. 30cm Höhe wurde mit einem Vorhang verkleidet. Um frühe Fußbodenreflexionen zu vermeiden, wurde der Linoleum-Fußboden plan mit schwerem Vorhangstoff und zusätzlich – mit einer Tiefe von 1,80m und über die gesamte Raumbreite symmetrisch zum Referenz-Abhörpunkt – mit einem weiteren Teppich bedeckt. Weiterhin wurden in den vorderen Ecken des Raumes zwei 2,10m hohe und 90cm breite, freistehende, breitbandig absorbierende Akustikelemente aufgestellt. Durch diese, die Akustik verändernden, Maßnahmen ließen sich stehende Wellen und Flatterechos vermeiden und die Auflagen der internationalen Richtlinien einhalten.

⁴⁰ EBU Tech.3276 – 2nd edition (5/1998).

⁴¹ ITU-R BS.775 und ITU-R BS.1116-1.

4 Versuchsbeschreibung

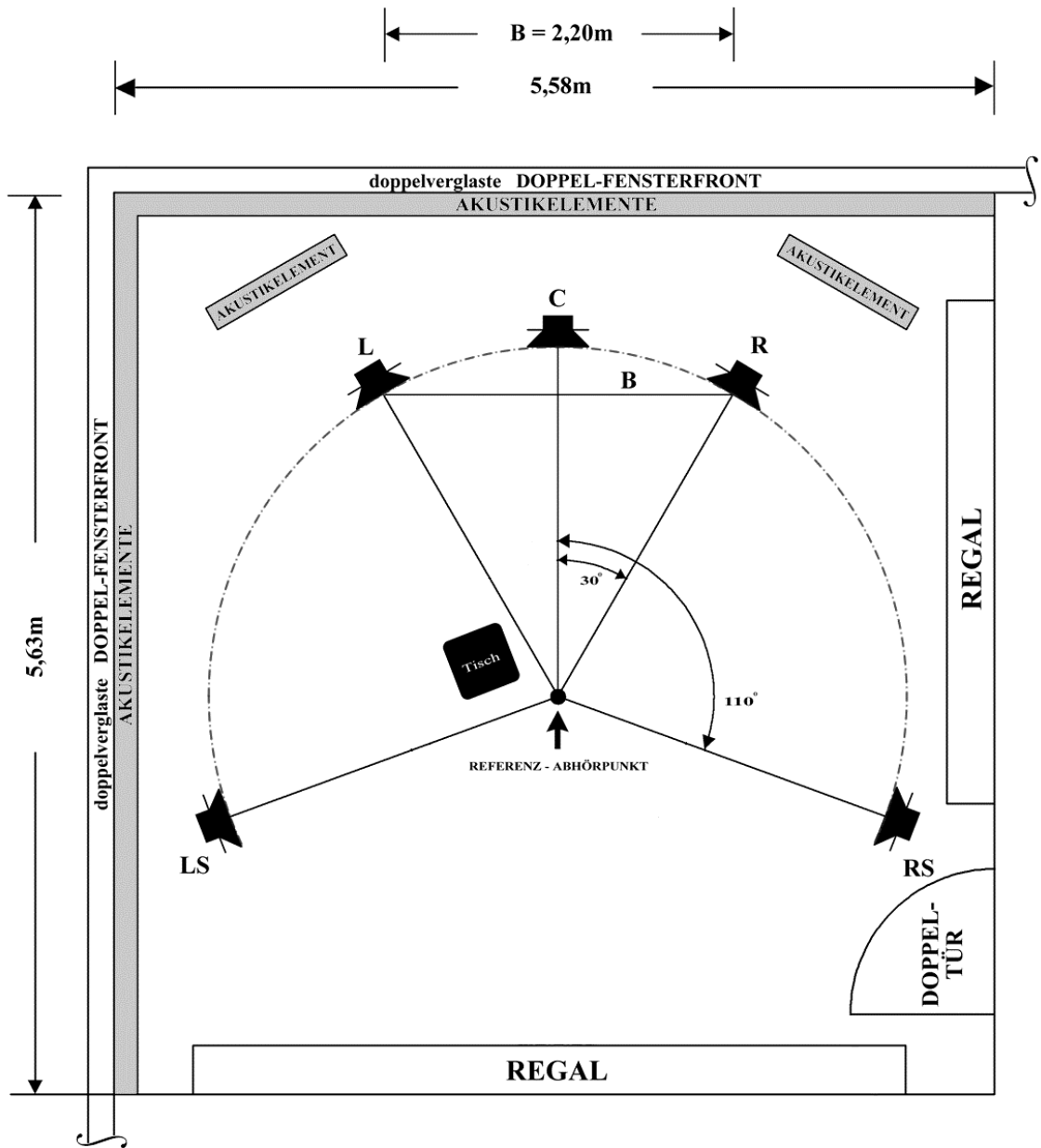
Entsprechend den Empfehlungen der ITU⁴² für die Lautsprecheranordnung von L, C, R, LS und RS bei Multi-Channel-Soundsystemen wurden die Manger-Lautsprechersysteme auf einem Kreis angeordnet. Dabei betrug die Basisbreite B zwischen L und R – sowie daraus folgend ebenso der Kreisradius – 2,20m. Die Lautsprechersysteme LS und RS wurden in einem 110°-Winkel zur Mittelsenkrechten (Referenzabhörpunkt – Center) äquidistant positioniert. Aus dieser Anordnung ergab sich ein Abstand von ca. 1m zu den umgebenden Wänden. Die Lautsprecher – für L und R kamen Aktivsysteme (Manger MSS-Studiomonitore), für C, LS und RS Passivsysteme (Manger Zerobox 109) mit dem gleichen Schallwandler der Firma Manger zum Einsatz – wurden auf eine einheitliche Höhe von 1,27m gebracht und mittels Betonplatten vom Fußboden entkoppelt (alle Angaben beziehen sich auf den Mittelpunkt der Schallwandlermembran).

Weiterhin sei erwähnt, dass links vor dem Probanden ein niedriger Tisch (40x40x35cm) stand, der neben dem Stax-Verstärker und Stax-Kopfhörer auch das Tablett mit den Steuerungsmodulen für die ABX-Software enthielt und auf dessen Oberplatte ein TFT-Monitor installiert war. Dieser Tisch beeinflusste mit einer Gesamthöhe von ca. 80cm den Direktschall nicht.

Der folgenden Grafik Nr.21 können der Grundriss, die Möblierung und die Lautsprecheranordnung entnommen werden:

⁴² ITU-R BS.1116-1 1997: 21.

4 Versuchsbeschreibung



Grafik Nr.21: Schematische Darstellung des Hörraums.

4 Versuchsbeschreibung

4.2.2 Messungen im Hörraum

4.2.2.1 Nachhallzeit

Der Nachhall wird durch die Reflexionen der umgebenden Wände des Hörraums bestimmt, die den Hörer mehr als 15ms nach dem Direktschall erreichen. Um wahrnehmbare akustische Effekte wie z.B. Flatterechos zu vermeiden, sollte er möglichst diffus sein.

Die nominale Nachhallzeit T_m eines Hörraumes, also das arithmetische Mittel der mit einem Terzbandfilter gemessenen (frequenzabhängigen) Nachhallwerte über einen Frequenzbereich von 200Hz bis 4kHz, sollte in einem Bereich von

$$0,2 < T_m < 0,4 \text{ s}$$

liegen. Hierbei ist die Nachhallzeit, als eine natürliche Klangeigenschaft eines Raumes, in Abhängigkeit zum Raumvolumen zu sehen und wird laut EBU⁴³ und ITU⁴⁴ wie folgt als Richtwert definiert:

$$T_m = 0,25 * \left(\frac{V}{V_0} \right)^{\frac{1}{3}} \text{ s}$$

wobei gilt: V = Raumvolumen des Hörraumes (in m^3)

V_0 = Referenz-Raumvolumen von 100m^3

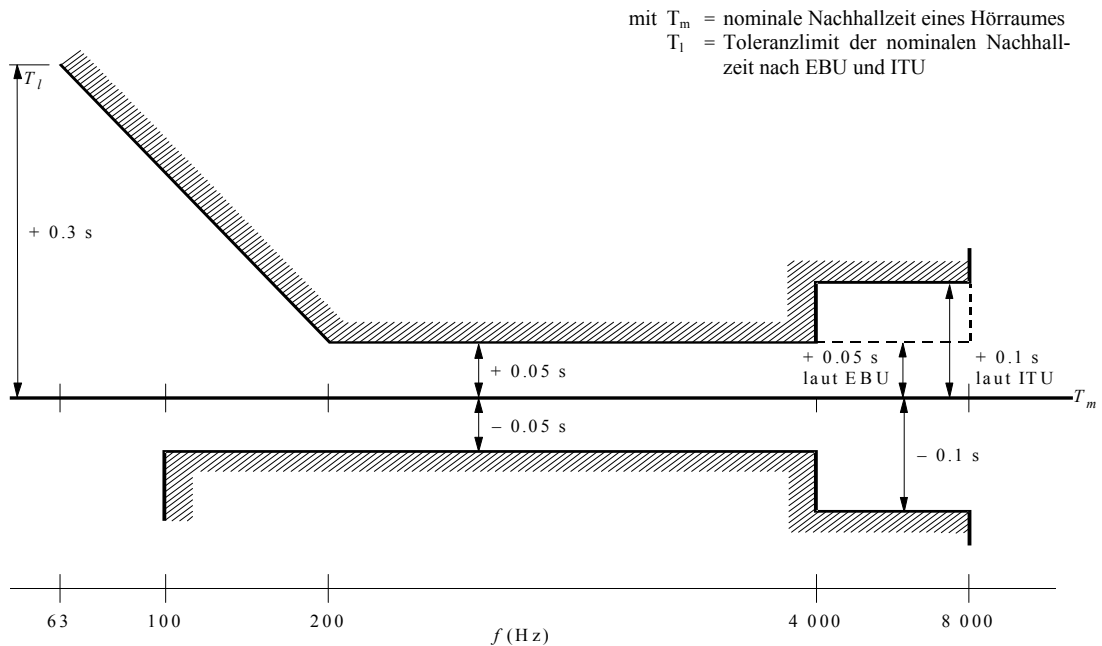
Aus dieser Formel errechnet sich für den Test-Hörraum eine nominale Nachhallzeit T_m von 0,238s.

Sowohl EBU als auch ITU geben für die mit Terzbandfilterung im Bereich zwischen 63Hz und 8kHz zu messende Nachhallzeit T ein Toleranzlimit an, wobei sich dieses geringfügig zwischen 4 und 8kHz unterscheidet: Während die ITU ein Limit von $\pm 0,1\text{s}$ angibt, toleriert die EBU in demselben Bereich lediglich einen Spielraum von $-0,1$ bis $+0,05\text{s}$, wie aus der folgenden Grafik Nr.22 ersichtlich wird:

⁴³ EBU Tech. 3276 – 2nd edition 1998: 6.

⁴⁴ ITU-R BS. 1116-1 1997: 17.

4 Versuchsbeschreibung



Grafik Nr.22: Toleranzlimit der nominalen Nachhallzeit nach EBU (gestrichelte Linie) und ITU (durchgezogene Linie).⁴⁵

Aus der folgenden Tabelle werden die Nachhallzeit T , sowie die Abweichung ΔT von der nominalen Nachhallzeit $T_m = 0,238$ s (nach EBU-Norm) des Hörraumes ersichtlich (gemessen am Referenz-Abhörplatz in einer Höhe von 1,20m):

Frequenz in Hz	60	80	100	125	160	200	250	315
Nachhallzeit in s	0,510	0,370	0,390	0,220	0,240	0,250	0,230	0,280
ΔT in s	+0,272	+0,132	+0,152	-0,018	+0,002	+0,012	-0,008	+0,042
Frequenz in Hz	400	500	630	800	1k	1,25k	1,6k	2k
Nachhallzeit in s	0,270	0,230	0,240	0,240	0,230	0,200	0,230	0,230
ΔT in s	+0,032	-0,008	+0,002	+0,002	-0,008	-0,038	-0,008	-0,008
Frequenz in Hz	2,5	3,15k	4k	5k	6,3k	8k		
Nachhallzeit in s	0,200	0,220	0,210	0,210	0,200	0,190		
ΔT in s	-0,038	-0,018	-0,028	-0,028	-0,038	-0,048		

Grafik Nr.23: Nachhallzeiten des Hörraumes.

Die Nachhallzeit-Messungen wurden mit dem Messgerät B&K 2231, der Kapsel B&K 4155 und dem aufsteckbaren B&K Terz- und Oktavfilter-Modul, Typ 1625, durchgeführt.

⁴⁵ In Anlehnung an ITU-R BS. 1116-1 1997: 17.

4 Versuchsbeschreibung

4.2.2.2 Hintergrundgeräusch

Laut EBU⁴⁶ und ITU⁴⁷ muss das kontinuierliche Hintergrundgeräusch, das durch Klimaanlage oder andere externe oder interne Quellen verursacht werden kann, im Hörbereich in einer Höhe von 1,20m über Fußbodenniveau gemessen werden. Hintergrundgeräusche impulsartiger, zyklischer oder tonaler Art sollten dabei auf keinen Fall wahrnehmbar sein.

Für empfohlene Werte geben sowohl EBU als auch ITU sogenannte Noise-Rating-Kurven an – dabei sollten möglichst die Werte der Noise-Rating-Curve NR 10 erreicht bzw. unterschritten, auf keinen Fall jedoch die der Noise-Rating-Curve NR 15 überschritten werden. Die im Hörraum erzielten Werte, sowie die der Noise-Rating-Curve NR 10 und 15 sind aus der nachfolgenden Tabelle und der Grafik Nr.25 ersichtlich. Die Messungen wurden mit dem Messgerät B&K 2231 und der Kapsel B&K 4155 durchgeführt:

Frequenz in Hz	31,5	63	125	250	500	1k	2k	4k	8k
NR 10	62,2	43,4	30,7	21,3	14,5	10,0	6,6	4,2	2,3
NR 15	65,6	47,3	35,0	25,9	19,4	15,0	11,7	9,3	7,4
Erreichte Werte	41,7	32,3	20,9	14,9	11,8	8,3	6,9	6,1	5,8

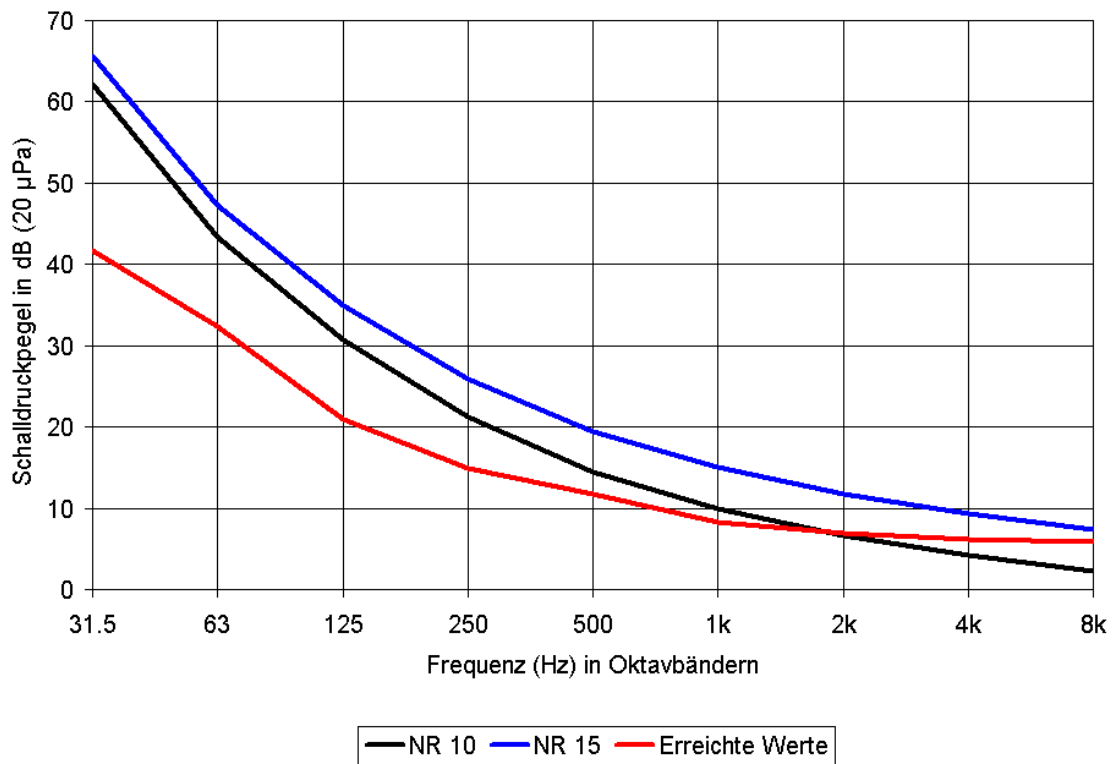
Grafik Nr.24: Hintergrundgeräusch in dB/SPL bezogen auf 20µPa nach EBU/ITU und erreichte Werte im Hörraum.

⁴⁶ EBU Tech. 3276 – 2nd edition 1998: 8.

⁴⁷ ITU-R BS. 1116-1 1997: 18.

4 Versuchsbeschreibung

Noise-Rating-Curve



Grafik Nr.25: Noise-Rating-Curves NR 10 und NR 15 und erreichte Werte im Hörraum.⁴⁸

4.2.2.3 Abhörpegel

Der Referenz-Abhörpegel $L_{LISTref}$ ist definiert als bevorzugter Abhörpegel, der mit Hilfe eines Rosa Rauschens bei -18dB (in Bezug auf dB Fullscale) am Referenz-Abhörpunkt – separat für jeden Kanal – eingemessen wird. Dies soll einen gleichen Schalldruckpegel in unterschiedlichen Hörräumen sicherstellen.

Laut EBU⁴⁹ und ITU⁵⁰ sollte die Verstärkung des Wiedergabekanals so eingestellt werden, dass sich folgender Schalldruckpegel (A-bewertet) am Referenzabhörpunkt ergibt:

$$L_{LISTref} = 85 - 10 \log(n) \text{ dB(A)}$$

wobei gilt: n = Anzahl der Wiedergabekanäle

⁴⁸ In Anlehnung an ITU-R BS. 1116-1 1997: 21.

⁴⁹ EBU Tech. 3276 – 2nd edition 1998: 8.

⁵⁰ ITU-R BS. 1116-1 1997: 18.

4 Versuchsbeschreibung

Bei fünf Wiedergabekanälen lässt sich ein Referenzschalldruckpegel $L_{LISTref}$ von 78,01dB(A) (gewählt 78dB(A)) berechnen. Dabei sollte der Unterschied zweier Kanäle zueinander nicht mehr als 1dB betragen. Aufgrund des nicht konstanten Schalldruckpegels bei Rosa Rauschen wurden die Kanäle wegen der Relevanz einer absoluten Pegelgleichheit zusätzlich mit einem 1kHz Sinuston bei -18dB eingemessen, wodurch eine Genauigkeit von 0,2 – 0,4 dB (0,1dB bei Sinuston) bei Rosa Rauschen erzielt werden konnte.

Gemessen wurde auch hier mit dem Messgerät B&K 2231 und der Kapsel B&K 4155.

4.3 Informationen zur verwendeten Technik

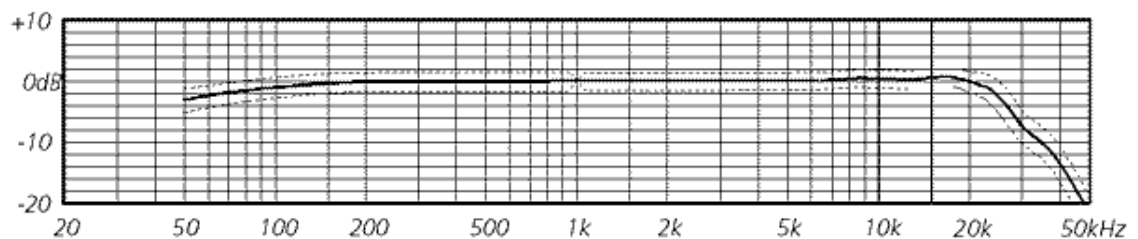
Um den Vorteil eines stark erweiterten Übertragungsspektrums hochauflösender digitaler Kodierungsverfahren nutzen zu können, sollte die zum Einsatz kommende Technik diesen Anforderungen ebenfalls gerecht werden. Im Folgenden soll die verwendete Technik – in der Reihenfolge des Signalflusses – kurz vorgestellt werden.

Das gesamte Equipment wurde, übereinstimmend mit den Spezifikationen, entsprechend lange vor Beginn der Tests eingeschaltet, so dass die Geräte eine stabile Arbeitstemperatur erreichen konnten.

- Mikrofone:

Bei der Wahl der Mikrofone sollte darauf geachtet werden, dass diese einen möglichst linearen, über die normale Audiobandbreite hinaus erweiterten Frequenzgang haben und des Weiteren eine möglichst kleine Membran besitzen, da die Impulsantwort umso präziser ist, je kleiner der Membrandurchmesser ist. Gewählt wurden Schoeps- und Sennheiser-Mikrofone:

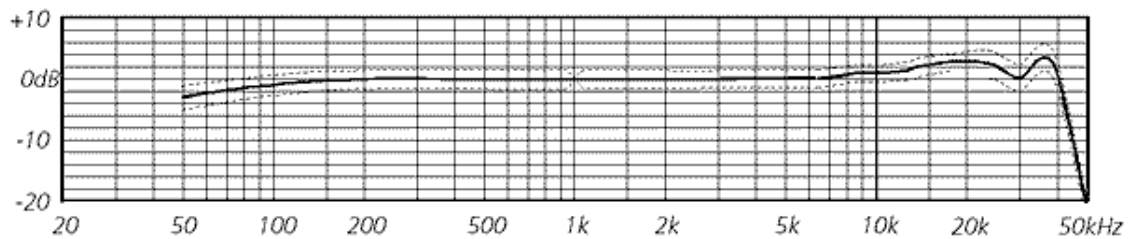
An Schoeps-Kapseln standen jeweils zwei MK2Sg und MK4g sowie eine MK41g des Colette-Systems zur Verfügung, die mit dem Mikrofonverstärker CMC 6-Ug xt („extended“) betrieben wurden. Die auf dem CMC 6 (s. Grafik Nr.26a) basierende Schaltung dieses Mikrofonverstärkers wurde von Schoeps so erweitert, dass der Übertragungsbereich/Höhenfrequenzgang sich bis über 40kHz fortsetzt (s. Grafik Nr.26b):



Grafik Nr.26a: Frequenzgang des CMC 6 – Standardausführung – mit Niere MK4.⁵¹

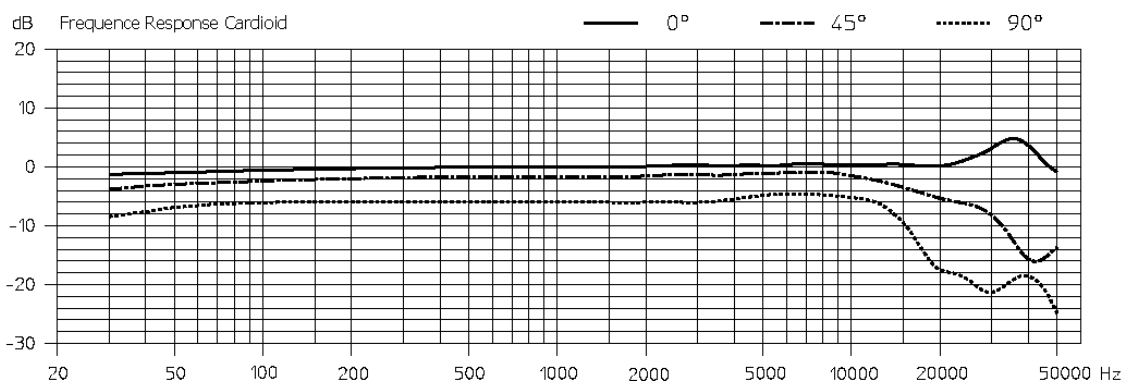
⁵¹ <http://www.schoeps.de> .

4 Versuchsbeschreibung



Grafik Nr.26b: Frequenzgang des CMC 6xt mit Niere MK4.⁵²

Das MKH800 von Sennheiser ist ein Studio-Richtmikrofon mit fünf schaltbaren Richtcharakteristiken – Kugel, breite Niere, Niere, Superniere und Acht. Es enthält einen Doppelwandler mit zwei im akustischen Verbund arbeitenden Gegentaktwandlern hoher Linearität; der Frequenzbereich ist, wie aus der nachfolgenden Grafik Nr.27 erkenntlich, bis 50kHz erweitert:



Grafik Nr.27: Frequenzgang Sennheiser MKH800 mit Nierencharakteristik.⁵³

- Mikrofonvorverstärker:

Der Lake People MIC-AMP F35/II ist ein zweikanaliger Mikrofon-Vorverstärker. Die Verstärkung erfolgt durch einen sogenannten Instrumentations-Verstärker in Form eines optimierten IC's⁵⁴, welches sich durch seine Rauscharmut bei hohen Verstärkungen und seine hohe Breitbandverstärkung (GBW = Gain Bandwidth Product) auszeichnet. Es ist – neben der Slew-Rate – entscheidend für die Güte eines Verstärkers und lässt sich aus dem erzielten Frequenzgang bei einer bestimmten Verstärkung berechnen. Der MIC-AMP F35/II hat bei einer 1000-fachen

⁵² Ebd.

⁵³ <http://www.sennheiser.com>.

⁵⁴ Eingesetzt wird der SSM 2017 von PMI. Dieser Instrumentations-Verstärker hat bei +60dB Verstärkung und 70kHz eine Absenkung von 0,5dB und ist somit ideal, was Schnelligkeit und Frequenzgang betrifft.

4 Versuchsbeschreibung

Verstärkung (+60dB) einen internen linearen Frequenzgang von weit über 200kHz, woraus sich ein theoretisches GBW von 200MHz ergibt.⁵⁵

- A/D und D/A-Konverter:

Sowohl die dCS 904 A/D Converter 24/192, als auch die dCS 954/955 D/A Converter 24/192 sind zweikanalige Stand-alone-Konverter, die nicht nur alle PCM-Standard-Abtastraten von 32 bis 192kHz, sondern auch DSD als Kodierungsformat anbieten. Beide Wandlertypen arbeiten ohne das sogenannte Gain Ranging⁵⁶ mit einer diskreten 5-Bit Architektur bei 64-fachem Oversampling, die auf dem patentierten dCS Ring DAC basiert. Dieser zeichnet sich vor allem durch seine sehr hohe Linearität aus.

Im PCM-Modus bietet der dCS 904 neben vier durch den Benutzer wählbaren Anti-Aliasing-Filtern (für jede Samplingfrequenz) auch ein Noise-Shaping erster, dritter und neunter Ordnung. Durch einen Auto-Modus wird gewährleistet, dass der für die gewählte Ausgangs-Wortbreite geeignete Noise-Shaper benutzt wird. Die Ausgangs-Wortbreite lässt sich in Ein-Bit-Schritten zwischen 16 und 24 Bit einstellen. Für die Ausgabe von 88,2- oder 96kHz-Daten bietet dCS, abgesehen von den einzelnen SDIF-2- oder AES3-Schnittstellen, ebenso die Möglichkeit, das Dual AES- bzw. für 176,4- oder 192kHz-Daten das Quad AES-Interface zu nutzen (sogenannter Comprehensive Digital Output). Letzteres Interface wurde beim vorliegenden Versuchsaufbau verwendet: Es besteht aus vier mit 44,1kHz auflösenden AES3-Leitungen, wobei jeweils zwei Leitungen pro Audio-Kanal genutzt werden (Leitungsaufteilung nach geraden und ungeraden Samples).

Im DSD-Modus stehen dem Benutzer acht frei wählbare, unterschiedliche DSD-Noise-Shaper zur Verfügung, die eine Beeinflussung der spektralen Verteilung des Quantisierungsrauschens ermöglichen. Die Ausgabe von DSD kann einerseits über die SDIF-2- oder SDIF-3-Schnittstelle oder andererseits über ein Pseudo-AES3 Format erfolgen (Daten-Bitmapping). Hierfür werden die vier digitalen AES-Outputs

⁵⁵ Entnommen aus der Bedienungsanleitung MIC-AMP F35/II 2001.

⁵⁶ Gain Ranging bedeutet, dass zwei oder mehr Wandlerbausteine mit unterschiedlich justierten Arbeitspunkten oder unterschiedlichen Kennlinien zusammen betrieben werden. Abhängig vom jeweiligen Pegel wird das Analogsignal einem der Wandler zugewiesen. Anschließend werden die Ausgangswerte der Wandler auf digitaler Ebene wieder addiert. Dieses Prinzip kann z.B. aufgrund der unterschiedlichen Wandlerkennlinien zu wahrnehmbaren Artefakten führen, die laut dCS das erreichbare Maß an Transparenz verringern könnten.

4 Versuchsbeschreibung

(DSD4) benutzt. Eine Erweiterung bildet der P3D-Modus, bei dem zwei Kanäle DSD auf drei AES3-Leitungen (44,1kHz/24Bit) ausgegeben werden.

Somit kann unter Verwendung des Comprehensive PCM- und P3D-Modus sowohl DSD als auch hochauflösendes PCM auf jedem herkömmlichen PCM-Recorder aufgenommen werden, der in der Lage ist, acht Kanäle mit 44,1kHz und 24Bit zu speichern.

Der dCS 954/955 bietet im PCM-Modus die Option, bei jeder Samplingfrequenz zwischen vier sogenannten „Anti-Imaging-Filtern“ zu wechseln. Diese haben unterschiedliche Übergangs-Bandbreiten, woraus sich demzufolge eine unterschiedlich präzise Impulsantwort ergibt (s. auch Kap.2.3). DSD-seitig kann zwischen vier Digitalfiltern gewählt werden, die durch eine unterschiedliche Ausgangssignal-Bandbreite eine optimale Anpassung an die Monitorkette garantieren. Selbstverständlich sind die Wandler in der Lage, alle digitalen Ausgangsformate des dCS A/D-Konverters zu verarbeiten.

Eine Fernbedienungs-Software ermöglicht, auf die Menüfunktionen von bis zu sechs dCS-Konvertern über den COM-Port eines Pentium-PC's zuzugreifen.

- Audio-Workstation:

Als Audio-Rechner wurde eine leistungsfähige Workstation mit einem Raid0-System konzipiert, auf der Pyramix „Virtual-Studio-System“ von Merging installiert wurde. Als „Digital Audio Input/Output-PCI-Lösung“ wurden zwei Mykerinos Mothercards (32Bit Floating Point) gewählt, die aufgrund ihrer zweiteiligen Architektur mit jeweils einem Mykerinos AES-EBU Daughter Board ausgestattet wurden. Diese bieten je 24 I/O-Kanäle über 12 AES-EBU I/O Leitungen, die über drei Brackets und Breakout-Kabel herausgeführt werden. Somit standen insgesamt 48 Audiokanäle mit 44,1kHz/24Bit über 24 AES-EBU I/O-Leitungen zur Verfügung.

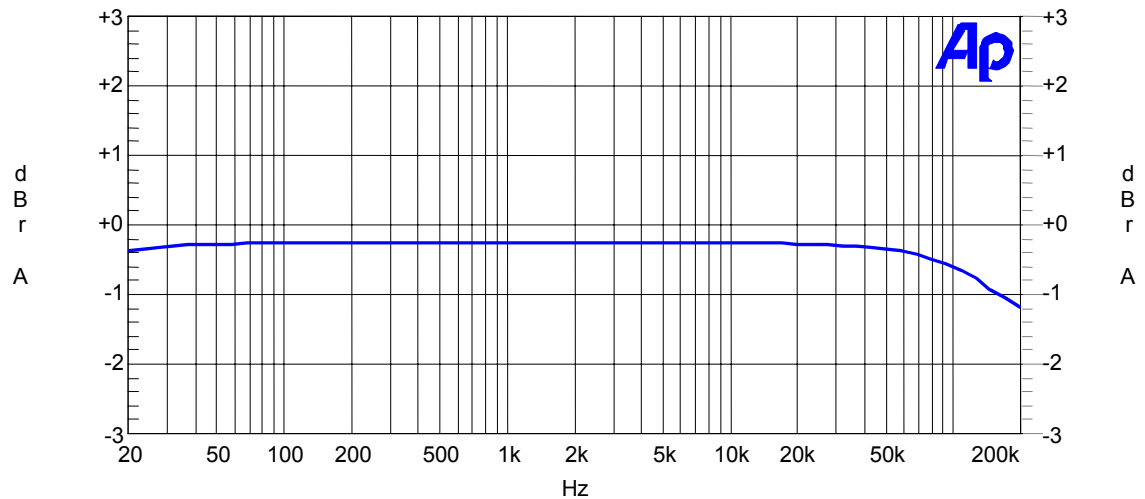
- Monitorunit und ABX-Software:

Sowohl die analoge Surround-Monitorunit MU 2000, als auch die ABX-Software basieren auf einer Eigenentwicklung der Emil-Berliner-Studios. Die Monitorunit mit symmetrischen Ein- und Ausgängen ermöglicht, zwei separate Stereo- und/oder Surround-Eingänge mittels der ABX-Software auf einen Ausgang zu routen. Dabei

4 Versuchsbeschreibung

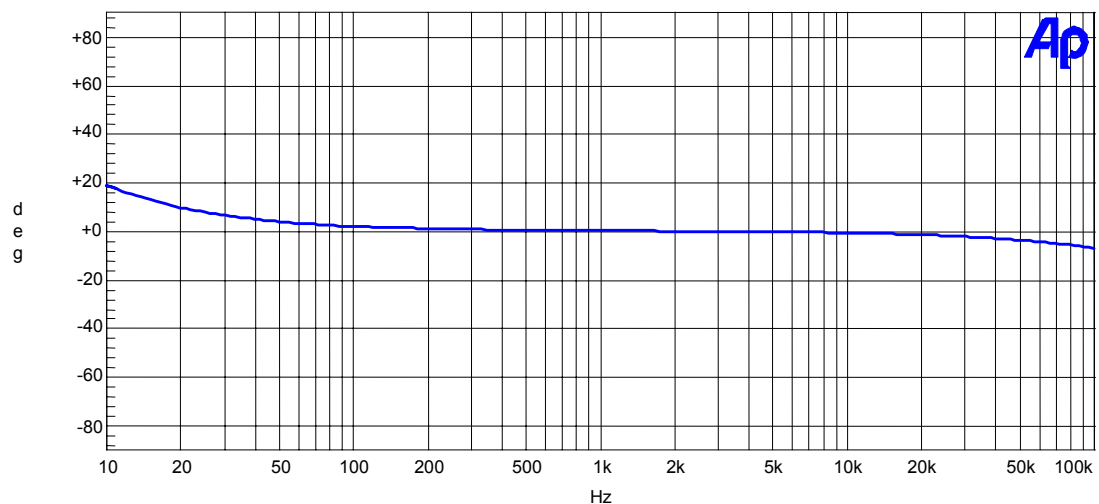
sind sowohl die Eingangs- als auch die Ausgangspegel abgleichbar. Der Frequenz- und Phasengang kann den folgenden Grafiken Nr. 28a und 28b entnommen werden:

Emil Berliner Studios



Grafik Nr.28a: Frequenzgang MU 2000 der Emil-Berliner-Studios.⁵⁷

Emil Berliner Studios



Grafik Nr.28b: Phasengang MU 2000 der Emil-Berliner-Studios.⁵⁸

- Endstufen:

Da im Gegensatz zu den aktiven Systemen (L und R) für C, LS und RS passive Lautsprecher verwendet wurden, waren für diese zusätzlich externe Endstufen not-

⁵⁷ Technische Unterlagen Emil-Berliner-Studios.

⁵⁸ Ebd.

4 Versuchsbeschreibung

wendig, die uns mit den AVM Monoblöcken „Evolution M3“ als zu den Lautsprechersystemen passende Verstärker ebenfalls durch die Firma Manger zur Verfügung gestellt wurden. Die Endstufen zeichnen sich besonders durch eine breitbandige Verstärkerschaltung aus, welche die exakte Reproduktion aller NF-Signale bis zu den im Ultraschallbereich liegenden Oberwellen ermöglicht. Dabei wird die Bandbreite von 700kHz sowohl unter Last als auch im Leerlauf erreicht. Die Verstärkerschaltung ist an allen vorkommenden Lastimpedanzen bis unter $0,5\Omega$ stabil. Durch diese Eigenschaften erzielen die Endstufen eine phasentreue und schnelle Wiedergabe des Audiosignals.

- Lautsprecher:

Im direkten Zusammenhang mit den Endstufen stehen die für einen Hörvergleich verwendeten Lautsprecher: An diese wird nicht nur der Anspruch eines erweiterten Frequenzbereichs, sondern ebenso der einer guten Impulswiedergabe gestellt. Werden Transienten, kurze Druckänderungen verursacht durch kurze Geräusche oder Impulse, nicht korrekt wiedergegeben, wird dadurch die Lokalisation bzw. die räumliche Wiedergabe beeinträchtigt. Diesbezüglich lassen sich mit Hilfe von sogenannten Sprungantwort-Tests etliche Aussagen treffen und Probleme vieler herkömmlicher Lautsprecher feststellen.

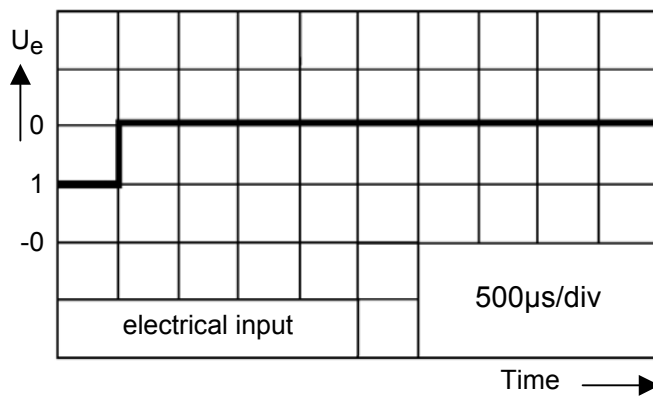
Aufgrund der besonderen Relevanz der Abhörsysteme für den Hörvergleich soll im Folgenden das Verfahren eines Sprungantwort-Tests ausführlich dargestellt werden:⁵⁹

Da die Luft einen Druckstoß nicht aufrecht erhalten kann, wird sich der aufgebaute Druck wieder einem statischen Wert, dem atmosphärischen Luftdruck, annähern. Dieser Vorgang verläuft zunächst über eine Unterdruckauslenkung, bevor sich der Luftdruck sukzessive dem vorherigen statischen Wert asymptotisch angleicht.

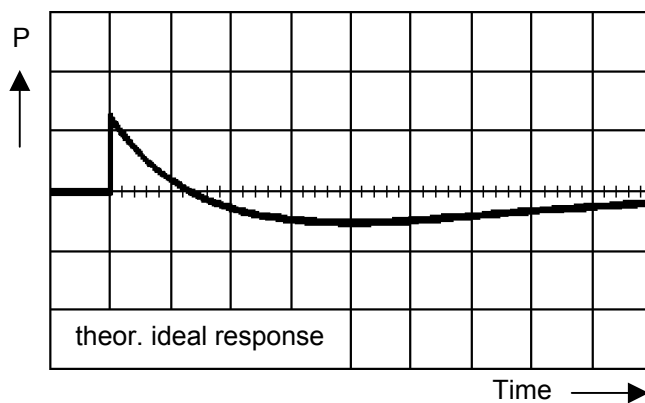
Wird ein elektrisches Eingangssignal gemäß Grafik Nr.29 auf einen Lautsprecher gegeben, so erhält man die in Grafik Nr.30 dargestellte theoretische Impulsantwort der Luft:

⁵⁹ Nach J.W.Manger 2000: Die akustische Realität: 8.

4 Versuchsbeschreibung



Grafik Nr.29: Elektrisches Eingangssignal.⁶⁰



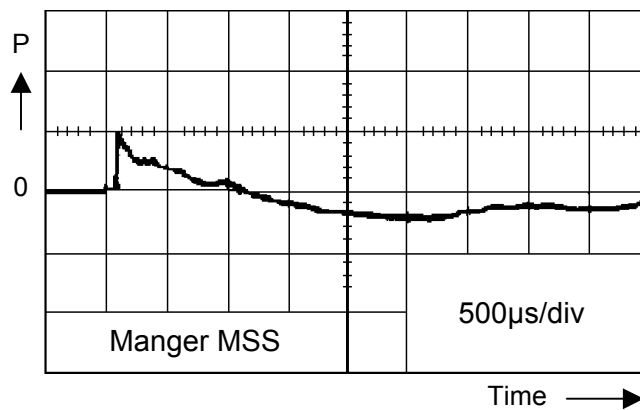
Grafik Nr.30: Theoretisch ideale Impulsantwort der Luft.⁶¹

Bei konventionellen Mehrwege-Lautsprechern werden mehrere in unterschiedlichen Frequenzbereichen arbeitende Treiber eingesetzt, die konstruktionsbedingt durch Masse-Feder-Schwingungssysteme zu unerwünschten Einschwingfehlern und daraus resultierenden Einschwinggeräuschen führen. Aufgrund der aufeinanderfolgenden Einschwingvorgänge der einzelnen Wege kommt es zu einem zeitversetzten Eigenschwingen, worin eine Ursache für lautsprechereigene Färbungen begründet liegt. Das Resultat ist zum einen neben Hörermüdigungserscheinungen eine leichtere Ortbarkeit der Lautsprecher, zum anderen eine deutlichere Fixierung auf den Referenz-Abhörpunkt, da dort die Laufzeit des Einschwingfehlers identisch ist. In den Grafiken Nr.31 und Nr.32 ist deutlich gegenübergestellt, wie sich dieser Fehler – im Gegensatz zum Manger Breitband-Schallwandler (MSW) – bei einem herkömmlichen Mehrwege-Lautsprechersystem bemerkbar macht:

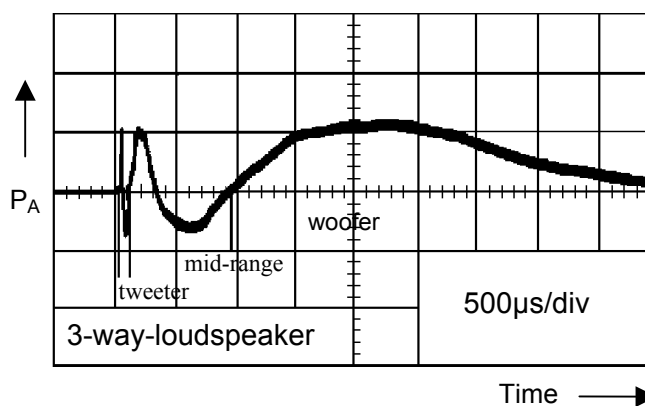
⁶⁰ <http://www.manger-msw.com> .

⁶¹ Ebd.

4 Versuchsbeschreibung

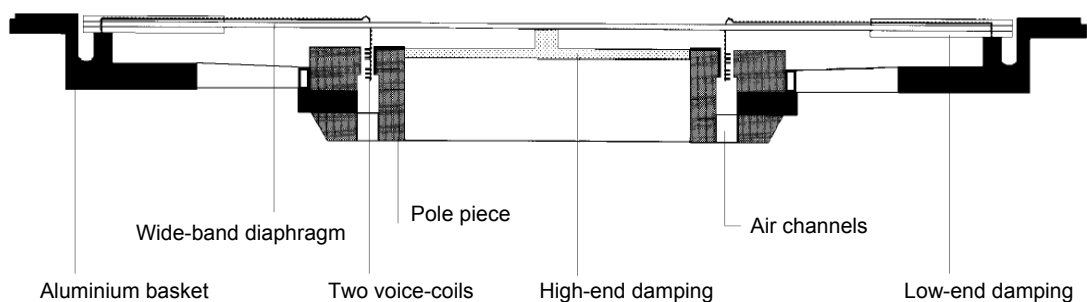


Grafik Nr.31: Impulsantwort Manger Schallwandler (MSW).⁶²



Grafik Nr.32: Impulsantwort eines Drei-Wege-Systems.⁶³

Der MSW wurde bei gleichzeitiger Erfüllung eines gleichmäßigen Amplituden-Frequenzgangs ebenfalls auf präzises Zeitverhalten hin konstruiert. Die Membran besteht aus einem speziellen, flexiblen Material, das Biegewellen ermöglicht, die sich – angeregt durch eine Schwingspule – radial unterschiedlich schnell nach außen ausbreiten (entspricht dem Ausbreitungsprinzip der Biegewellen auf der Basilar-membran im Innenohr):



Grafik Nr.33: Skizze des MSW.⁶⁴

⁶² Ebd.

⁶³ Ebd.

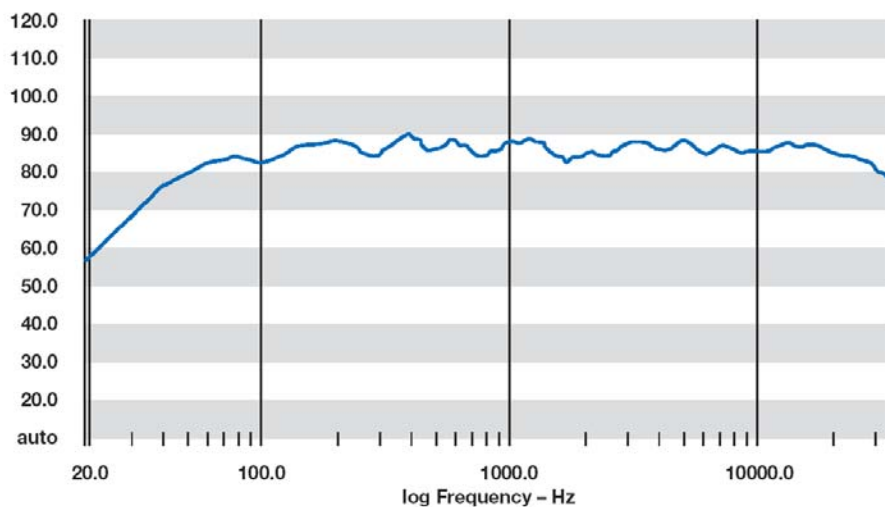
⁶⁴ Ebd.

4 Versuchsbeschreibung

Weitere Gründe für die Auswahl der Manger-Systeme waren die extrem schnelle Anstiegszeit von 13 Mikrosekunden sowie der erweiterte Frequenzbereich bis 35kHz des Manger Schallwandlers.

Wie bereits erwähnt wurden bei dem Hörvergleich unterschiedliche Lautsprecher-Systeme verwendet: Für C, LS und RS kam die Manger Zerobox 109, für L und R das aktive Manger Schallsystem (MSS) zum Einsatz. Letztere bestehen aus dem TT-Submodul und dem MSW-Topmodul, das mit zwei Endstufen in Brückenschaltung mit einer Anstiegszeit von $1,5\mu\text{s}$ (entspricht einer Bandbreite von ca. 330kHz) arbeitet. Dieser Studiomonitor ist mit drei MSW's ausgestattet, wobei die jeweils seitlich angebrachten MSW's pneumatisch eine unendliche Schallwand erzeugen. Die Zerobox 109 ist eine 2-Wege-Kompaktbox mit geschlossenem Gehäuse bei einer Trennfrequenz von 140Hz.

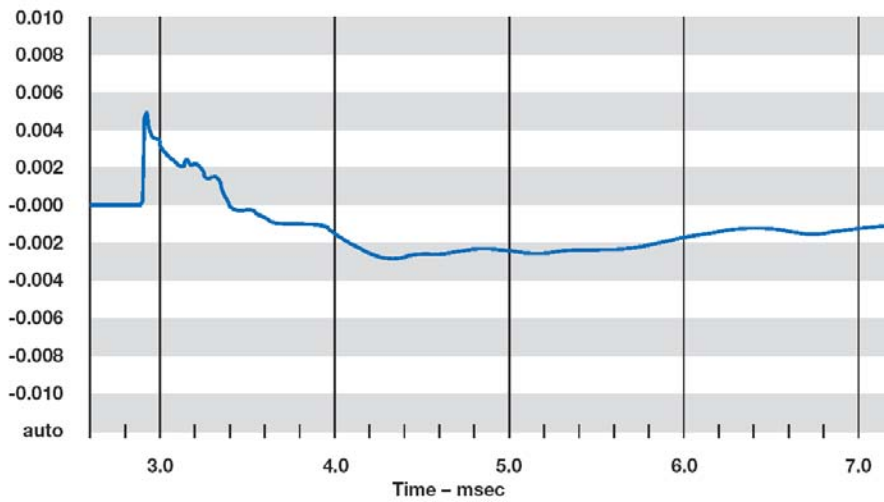
In den Grafiken Nr.34a und 34b, wie auch Nr.35a und 35b folgen jeweils Frequenzgang- und Sprungantwort-Messungen der verwendeten Lautsprechersysteme Manger Zerobox 109 sowie Manger MSS-Studiomonitor:



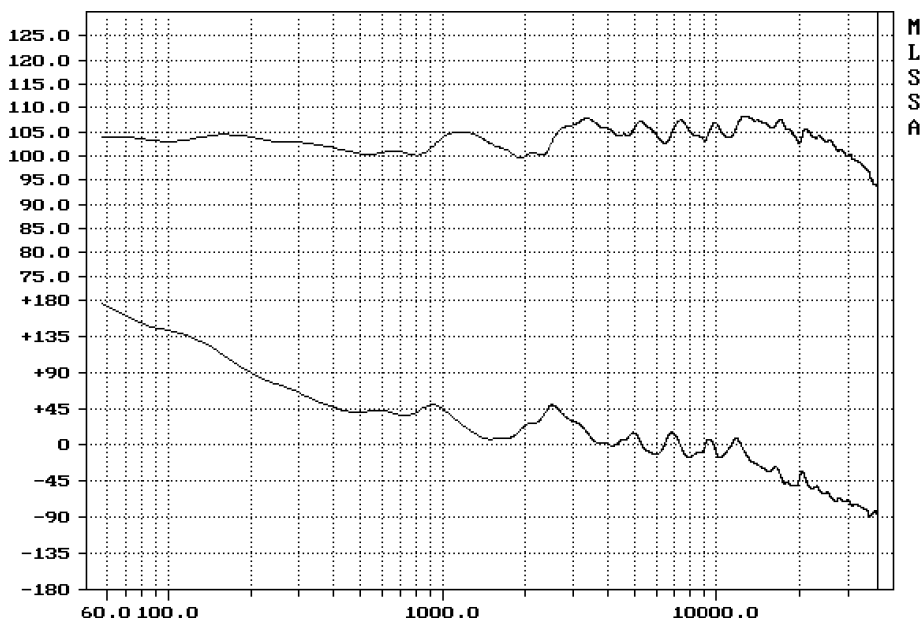
Grafik Nr.34a: Frequenzgang Zerobox 109.⁶⁵

⁶⁵ Ebd.

4 Versuchsbeschreibung



Grafik Nr.34b: Sprungantwort Zerobox 109.⁶⁶

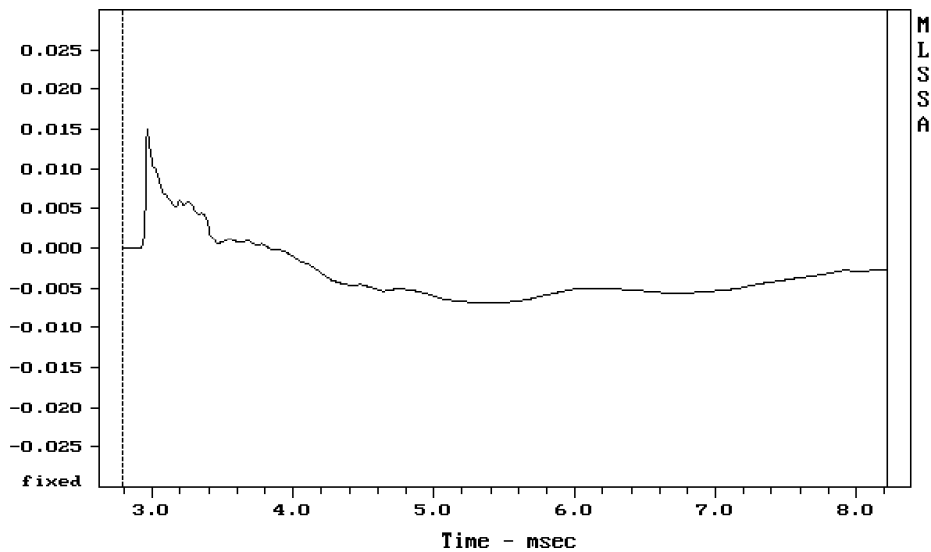


Grafik Nr.35a: Frequenz- und Phasengang MSS-Studiomonitor.⁶⁷

⁶⁶ Ebd.

⁶⁷ Technische Unterlagen Manger Products.

4 Versuchsbeschreibung



Grafik Nr.35b: Sprungantwort MSS-Studiomonitor.⁶⁸

- Kopfhörer:

Als Kopfhörer wurde der hochwertige und bereits als Studio-Standard geltende Stax SR-λ Professional mit dem Röhrenverstärker SRM-006t eingesetzt. Der elektrostatische Kopfhörer überträgt einen Frequenzbereich von 8Hz bis 35kHz, der Röhrenverstärker reicht nach Angaben des Herstellers von DC bis ca. 44kHz/+0, -1,5dB.

- Audiokabel:

Die gesamte Verkabelung wurde ausnahmslos mit neuen, hochwertigen Analog- und Digitalkabeln der Firma Klotz ausgeführt:

Als analoge Audiokabel wurden ausschließlich solche der Serie M1 verwendet, die auf dem flexiblen und mit 60pF/m (Leiter/Leiter) sehr niederkapazitiven Mikrofonkabel MC 2000 basiert. Sämtliche digitalen 110Ω-Audiokabel sind AES-EBU-Patchkabel der Serie AY vom Kabeltyp AEY 122.

Des Weiteren wurden Wordclockkabel der Serie SWC mit dem Kabel V06L37 und Y-Adapter vom Typ PYB5 benutzt.

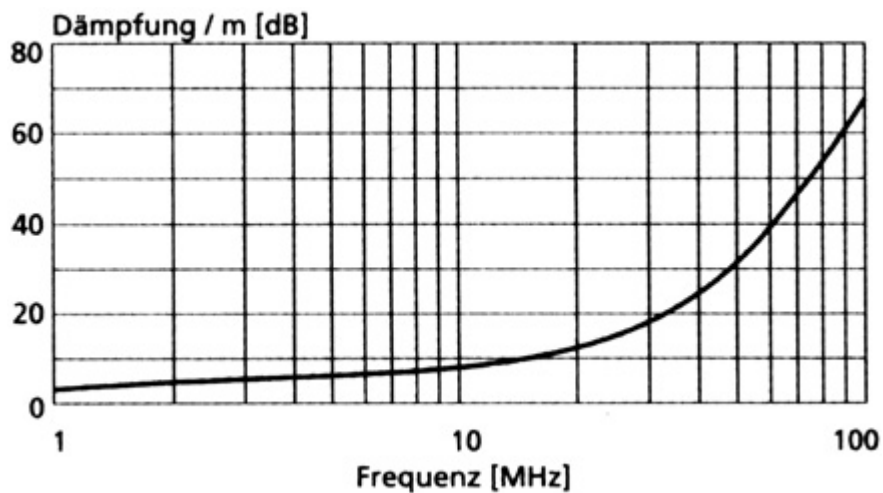
- Netzfilterung:

Um neben dem geregelten Netz des Erich-Thienhaus-Instituts noch zusätzlich eine Netzfilterung mit einzubeziehen, wurde die gesamte Technik über ein ferromag-

⁶⁸ Ebd.

4 Versuchsbeschreibung

Netzfilter-Kabel, das sich aus 7m Kabel und einer 8-fach Steckdosenleiste zusammensetzt, angeschlossen. Dieses eliminiert auf passive Weise breitbandige HF-Störungen: Die drei Kupferleiter sind jeweils mit einer flexiblen Ferritmasse sehr hoher Permeabilität umgossen, in die durch die Störsignale elektrische Wirbelströme induziert werden, welche jedoch, bedingt durch die hohe Impedanz, innerhalb des Kabels sehr effektiv absorbiert werden. So wird eine Dämpfung von 5 – 70dB/m (je nach Frequenz) erreicht:



Grafik Nr.36: Dämpfung des ferromag-Netzfilter-Kabels in Abhängigkeit der Frequenz.⁶⁹

Um die Dämpfung in einem breiten Spektralbereich noch zu verbessern, wird das Kabel zusätzlich einem gerichteten Magnetfeld ausgesetzt, wodurch sich die Polarisation der Ferrite noch erhöht.

⁶⁹ <http://www.ferromag.de> .

4.4 Versuchsablauf

Der **ABX**-Test ist ein Testverfahren, das ermöglicht, zwei unterschiedliche Audiosignale **A** und **B** „blind“ miteinander zu vergleichen. Ziel des Testverfahrens ist, das Signal **X**, welches von einer ABX-Software mit Hilfe eines Zufallsgenerators nach jeder Entscheidung neu auf **A** oder **B** zugewiesen wird, mit den Originalen **A** und **B** klanglich zu vergleichen und anschließend dementsprechend zuzuordnen (**X=A** oder **X=B**). Zwischen den drei parallel ablaufenden Musikbeispielen **A**, **B** und **X** hat der Proband die Möglichkeit, frei und beliebig oft zu wechseln, so dass er keinem zeitlichen Entscheidungsdruck ausgesetzt ist. Um eine signifikante Aussage treffen und somit Zufallsentscheidungen in aller Regel ausschließen zu können, muss der Proband mindestens 16 dieser Entscheidungen durchlaufen. Anhand des statistischen Auswertungsverfahrens kann anschließend eine Aussage getroffen werden, ob ein Unterschied wahrgenommen wurde oder nicht. Das Testverfahren des Doppelblindtests wird von der ITU⁷⁰ für die Durchführung solcher Hörvergleiche empfohlen.

Im Fall des vorliegenden Hörvergleichs war **A** entweder DSD und **B** komplementär dazu PCM (176,4kHz/24Bit) oder umgekehrt dazu **A** = PCM und **B** = DSD⁷¹. An Stelle von 16 musste der Proband aufgrund der höheren Aussagekraft⁷² 20 dieser Entscheidungen durchlaufen.

Jeder Hörvergleich besitzt ein subjektives Moment – den Probanden. Dieser stellt sich im Allgemeinen selber unter Leistungsdruck, obwohl im hier beschriebenen Hörvergleich die Intention nicht die Untersuchung des Hörvermögens der Testperson selbst war, sondern die Unterscheidbarkeit zweier Audiosignale getestet und beurteilt wurde. Um diesen Druck weitestgehend zu minimieren, wurde versucht, die Testsituation durch eine möglichst angenehme Atmosphäre im Hörraum zu entspannen. Aufgrund der Entscheidung, jedem Probanden die Wahl eines Musik- oder Klangbeispiels zu überlassen, ergab sich trotz des höheren zeitlichen Aufwands zwangsläufig, dass jeder Proband einzeln – gerade deswegen jedoch ungestört – getestet wurde. Dadurch wurde ihm ebenso ermöglicht, den Ablauf des Hörtests und die individuell für die 20 Ent-

⁷⁰ ITU-R BS.1116-1 1997: 7.

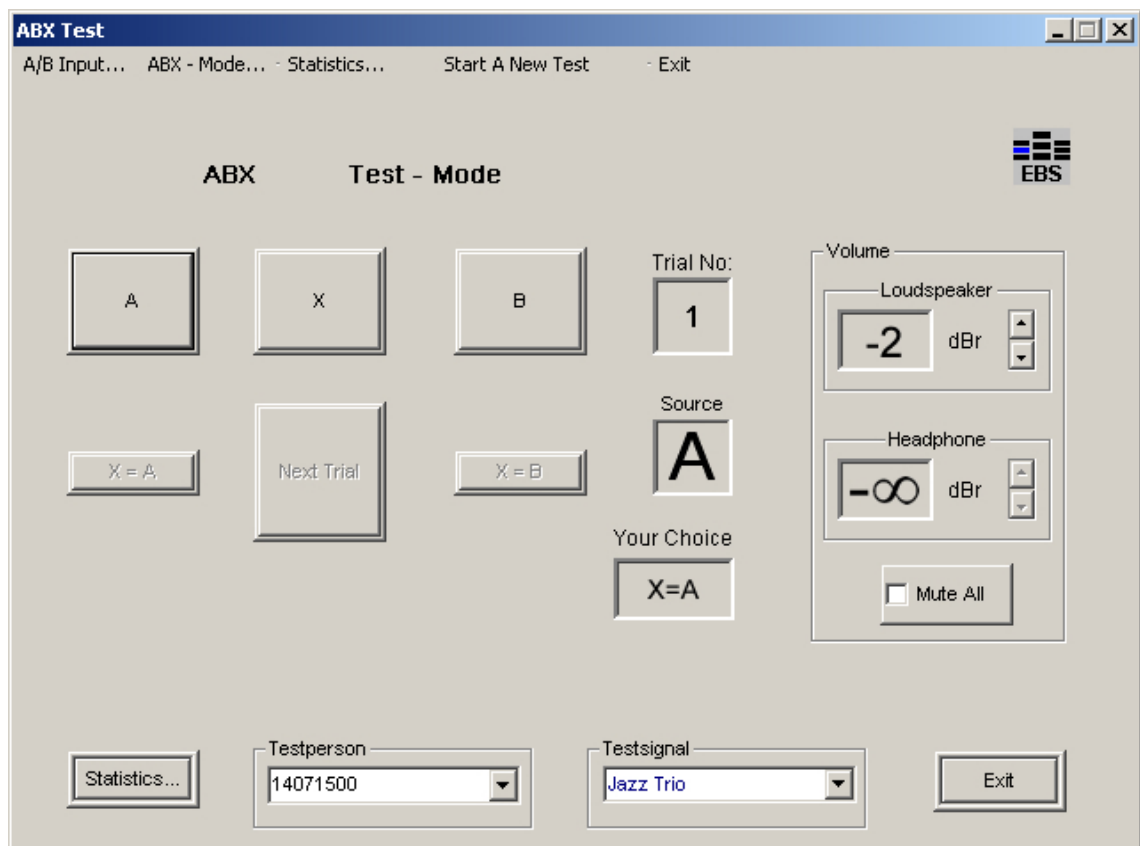
⁷¹ Die Zuordnung von A und B auf DSD oder PCM blieb selbstverständlich während eines Testdurchgangs pro Proband unverändert.

⁷² Siehe Kapitel 6 „Statistische Auswertungsmethode“.

4 Versuchsbeschreibung

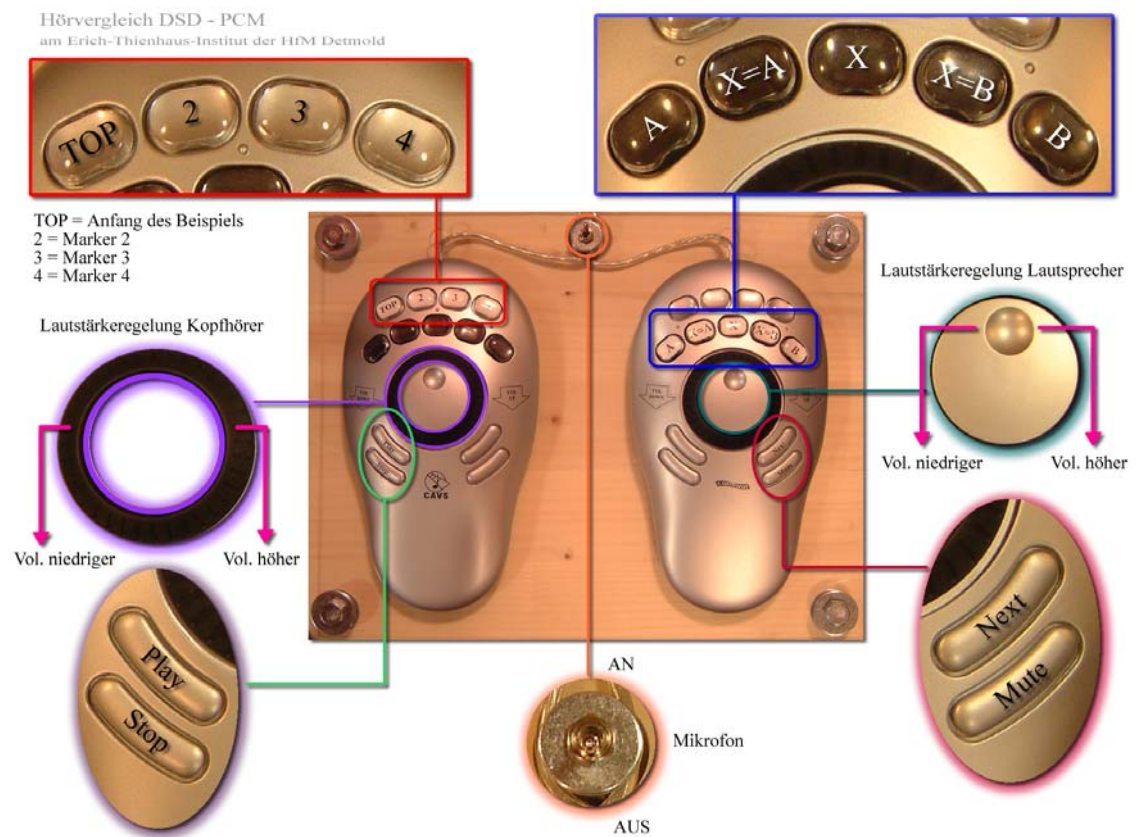
scheidungen benötigte Zeit selber zu bestimmen, sowie gegebenenfalls eine erforderliche Pause einzulegen.

Der Hörvergleich unterteilte sich in zwei Phasen: Zunächst gab es eine Lernphase, innerhalb derer sich der Proband – nach einer genauen Erklärung des Testablaufs und einer technischen Einweisung durch einen der beiden Testleiter – an die relativ einfache Bedienung der Steuerungsmodule für die ABX-Software wie auch die Software als solche gewöhnen und nach eigenem Interesse in das zur Verfügung gestellte Musikmaterial (Stereo und/oder Surround) hinein hören konnte. Somit war eine Simulation der „realen“ Testsituation möglich. Der Proband hatte die Möglichkeit, über definierte Markerpositionen gezielt musikalische Abschnitte wiederholt zu hören. Im Gegensatz zur späteren Testphase erhielt der Proband während der Lernphase nach einer getroffenen Entscheidung eine Bestätigung, ob er **X** richtig zugeordnet hatte oder nicht.



Grafik Nr.37: Bildschirmfenster der ABX-Software, die der Proband mit Hilfe eines Steuerungsmoduls (s.u.) bedienen konnte.

4 Versuchsbeschreibung



Grafik Nr.38: Steuerungsmodul für die Bedienung der ABX-Software: Verwendet wurden zwei auf einem Tablett montierte und modifizierte „Shuttle Pro“-Bedienungen von Contour.

Da es aus statistischen Gründen notwendig ist, dass die 20 Entscheidungen eines ABX-Tests mit jeweils nur einem Musikbeispiel bei gleicher Wiedergabeart durchgeführt werden, musste sich der Proband während der Lernphase für ein Musikbeispiel (Stereo oder Surround) und Wiedergabe über Lautsprecher oder Kopfhörer (Stereo) entscheiden. Um Ermüdungserscheinungen des Gehörs schon vor dem eigentlichen Testdurchlauf zu vermeiden, sollte diese Phase allerdings nicht länger als ca. 20-25 Minuten (einschließlich Erklärungen) dauern⁷³.

Nach einer optionalen kleinen Pause wurde die zweite Phase, der eigentliche Hörvergleich, mit dem zuvor gewählten Beispiel durchgeführt. Da die verwendete ABX-Software keine detaillierte Auflistung der einzelnen Entscheidungen der Probanden ausgibt, wurde nach 10 getroffenen Zuweisungen von X jeweils ein Zwischenergebnis vom Testleiter abgespeichert, das im Nachhinein eine Aussage über eine eventuelle Entwicklung während des Testverlaufs ermöglichen sollte. Dieses (wie auch

⁷³ Die Gesamtdauer des Hörvergleichs war aufgrund des äußerst individuellen Parameters der Entscheidungsgeschwindigkeit sehr verschieden, in der Regel wurden jedoch 60 Minuten für die Kombination von Lern- und Testphase nicht überschritten. Bezüglich des Ergebnisses ließ sich keine Korrelation zwischen Entscheidungstempo und Treffsicherheit feststellen.

4 Versuchsbeschreibung

das Gesamtergebnis) erfuhr der Proband erst nach dem Ausfüllen zweier Fragebögen, die er nach Absolvierung des Tests erhielt.

Da es sich beim vorliegenden Hörvergleich in erster Linie um einen Verifikationstest handelt, bei dem die auditive Differenzierung zweier unterschiedlicher Signale im Vordergrund steht, wurde auf detailliertere Fragen nach subjektiver klanglicher Bewertung oder Bevorzugung eines der beiden Audiosignale seitens der Probanden verzichtet.

Nach Beendigung des ABX-Tests bekamen alle Probanden den nachfolgenden Fragebogen Nr.1 ausgehändigt, auf dem eine persönliche Beurteilung des eben gehörten vermerkt werden konnte. Entschied sich ein Proband – nur nach relativ zügiger Durchführung der 20 Entscheidungen möglich – den ABX-Test unter identischen Bedingungen mit einem anderen Musikbeispiel zu wiederholen, musste – nach Abgabe von Fragebogen Nr.1 – dieser optionale weitere Test zunächst beendet werden. Erst nach dem zuletzt erfolgten ABX-Test erhielt der Proband vor Bekanntgabe des Ergebnisses Fragebogen Nr.2, der neben rein biographischen auch Fragen zu Musikhörergewohnheiten und -erfahrungen und allgemein dem Testablauf beinhaltete:

	HV# _____
Beispiel: _____	
<input type="checkbox"/> Stereo <input type="checkbox"/> Surround	
Haben Sie im Fall Stereo mit Kopfhörer gehört?	<input type="checkbox"/> ja <input type="checkbox"/> nein
Haben Sie einen Unterschied gehört?	<input type="checkbox"/> ja <input type="checkbox"/> nein
Wenn ja, können Sie diesen Unterschied verbalisieren?	
<div style="border: 1px solid black; height: 150px; width: 100%;"></div>	
Wie viel Hörerfahrung haben Sie mit der von Ihnen gewählten Besetzung?	_____

Grafik Nr.39: Fragebogen Nr.1.

Fragebogen zum Hörvergleich „DSD mit High Resolution PCM“

Dominik Blech und Min-Chi Yang

Persönliche Angaben:

HV# _____

Geburtsjahr: _____ Geschlecht: _____

Beruf: _____

Schule/Studiengang: _____

Spielen Sie ein Instrument, wenn ja, welches? ja nein _____

andere musikalische Aktivitäten: _____

Eigene Hörgewohnheit:

Welche Musikrichtung hören Sie bevorzugt? Bitte auch Mehrfachnennungen!

- Klassik
 - Orchesterwerke
 - Kammermusik
 - geistl. Musik
 - Oper
 - Lied
- Jazz
- Rock/ Pop
- Folklore

Wie viele Stunden hören Sie durchschnittlich pro Tag Musik? _____

Hören Sie eher kurze oder längere Einheiten? _____

In welcher Lautstärke hören Sie durchschnittlich (1=leise bis 10=sehr laut)? _____

Wie hören Sie Musik? nebenbei konzentriert Disko oder Kneipe
(Mehrfachnennungen) Konzert

Welches Medium hören Sie? LP Cassette MP3-Player Radio
(Mehrfachnennungen) CD DVD-A DVD-V SACD

Haben Sie Hörerfahrung mit Surround? Seit wann? ja nein _____

Haben Sie eine Surround-Anlage zu Hause? ja nein

Wenn ja, welche Art von Player? DVD-A DVD-V SACD

Bitte beachten Sie auch die Rückseite...

4 Versuchsbeschreibung

Verfassung vor dem Hörvergleich:

Wie lange hat Ihre Anfahrt gedauert? _____

Mit welchem Verkehrsmittel sind Sie gekommen? _____

Wie viel Zeit lag zwischen Anfahrt und Beginn des Hörvergleichs? _____

Haben Sie während der Anfahrt Musik/Radio gehört? _____

Haben Sie sich ausgeruht und konzentrationsfähig gefühlt? _____

Eindrücke während des Hörvergleichs:

Wie anstrengend fanden Sie den Hörvergleich? Können Sie Gründe dafür nennen?

Wie haben Sie die Atmosphäre im Hörraum empfunden?

Haben Sie Umgebungsgeräusche wahrgenommen? Wenn ja, welcher Art? Haben Sie diese als störend empfunden?

Hat Sie etwas an dem von Ihnen gewählten Musikbeispiel gestört, was hat eine Entscheidung eventuell erschwert?

WIR DANKEN IHNEN FÜR IHRE MITHILFE UND IHR INTERESSE!!!

Dominik Blech und Min-Chi Yang

5 Statistische Auswertungsmethode

Für die Auswertung der Ergebnisse des ABX-Tests ist zunächst die Wahl einer statistischen Methode notwendig, um anschließend mathematisch fundierte und interpretierbare Aussagen treffen zu können. Von der ABX-Company wird hier das statistische Modell der Binomialverteilung als sehr nützlich empfohlen.

5.1 Die Binomialverteilung

Bei einem Zufallsexperiment mit lediglich zwei möglichen Ausgängen gilt, dass das Ereignis A mit der Wahrscheinlichkeit p ($0 < p < 1$) und das zu A komplementäre Ereignis B mit der Wahrscheinlichkeit $q = 1 - p$ eintritt. Wird dieses Zufallsexperiment mehrmals hintereinander ausgeführt, so spricht man vom sogenannten Bernoulli-Prozeß. Voraussetzung hierfür ist, dass die aufeinanderfolgenden Experimente sich nicht gegenseitig beeinflussen, d.h., dass der Ausgang der i -ten Durchführung unabhängig vom Ausgang der j -ten Durchführung ist.

Die Wahrscheinlichkeiten p für das Ereignis A und q für das Ereignis B gelten auch für jede Wiederholung des Experiments. Das jeweilige Ereignis tritt also grundsätzlich mit der gleichen Wahrscheinlichkeit ein.

Finden bei einem Bernoulli-Prozeß n Versuche statt, so kann das Ereignis A unter den n beobachteten Ergebnissen 0- bis n -mal auftreten. Ist nun X die Anzahl der Ergebnisse A unter den n beobachteten Ergebnissen, dann liefert die Binomialverteilung die Wahrscheinlichkeit für das Ereignis ($X = x$).

Sei $BBA \dots AB$ eine Folge der Länge n , in der x -mal der Ausgang A und $(n-x)$ -mal der Ausgang B auftritt, so lässt sich aufgrund der Unabhängigkeit der einzelnen Versuche die Wahrscheinlichkeit $P(BBA \dots AB)$ wie folgt berechnen:

$$P(BBA \dots AB) = P(B)P(B)P(A) \dots P(A)P(B) = p^x(1-p)^{n-x}$$

Zwar sind bei diesem Versuch die Anzahl der Ereignisse A interessant, nicht aber deren konkrete Platzierung innerhalb der Abfolge, d.h., im wievielten Versuch ein Ereignis A

5 Statistische Auswertungsmethode

aufgetreten ist. Unter den n möglichen Positionen sind genau x Positionen mit dem Ereignis A zu besetzen, was auf $\binom{n}{x}$ verschiedene Möglichkeiten erfolgen kann. Hieraus

ergibt sich die gesuchte Wahrscheinlichkeit $P(X=x)$ der Binomialverteilung:

$$P(X=x) = p_x = \binom{n}{x} p^x q^{n-x} \quad \text{mit } x = 0, 1, \dots, n$$

$$= \binom{n}{x} p^x (1-p)^{n-x} \quad \text{und} \quad \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

p : Wahrscheinlichkeit für das Eintreten des Ereignisses A (beim Einzelversuch) mit $(0 < p < 1)$

q : Wahrscheinlichkeit für das Eintreten des zum Ereignis A komplementären Ereignisses B (beim Einzelversuch) mit $q = 1 - p$

n : Anzahl der durchgeführten Experimente/ Einzelversuche

x : Anzahl der Versuche, bei denen das gewünschte Ereignis eintritt

Bei solch einem Auswertungsverfahren soll geprüft werden, ob eine zuvor aufgestellte Hypothese H aufgrund der Ergebnisse des Tests beibehalten werden kann oder ob man sie ablehnt und sich damit für die sogenannte Gegenhypothese G entscheidet.

Bei Entscheidungen, die auf der Grundlage von Tests getroffen werden, können grundsätzlich jedoch auch Fehlentscheidungen bezüglich der Interpretation der Ergebnisse auftreten:

- Entscheidung für G , obwohl H richtig ist – dieser Fehler wird als Fehler erster Art bezeichnet
- Entscheidung für H , obwohl G richtig ist – dieser Fehler wird als Fehler zweiter Art bezeichnet

5 Statistische Auswertungsmethode

Die Möglichkeiten, sich richtig oder falsch zu entscheiden, lassen sich aus untenstehender Tabelle ersehen:

	H trifft zu (Realität)	G trifft zu (Realität)
Entscheidung für H	richtige Entscheidung	Fehler zweiter Art
Entscheidung für G	Fehler erster Art	richtige Entscheidung

Es lässt sich zeigen, dass die Wahrscheinlichkeit für einen Fehler erster Art umso kleiner ist, je größer die Wahrscheinlichkeit für einen Fehler zweiter Art ist. Man kann also nicht bei einer vorgegebenen Anzahl von Einzelversuchen die Wahrscheinlichkeiten für beide Fehlermöglichkeiten klein halten! In der Praxis geht man daher so vor, dass man die Wahrscheinlichkeit für den Fehler erster Art vorgibt, um zumindest eine Fehlerart unter Kontrolle halten zu können.

Die Wahrscheinlichkeit für den Fehler erster Art bezeichnet man auch als Signifikanzniveau oder Irrtumswahrscheinlichkeit, da durch ihn – durch Konvention – bestimmt wird, ab wann ein Ergebnis eine signifikante Abweichung zu dem unter der Hypothese erwarteten Ergebnis zeigt. Das Signifikanzniveau gibt somit an, ab welcher Abweichung man nicht mehr bereit ist, diese als zufällig zu bezeichnen. Die kritische Wahrscheinlichkeit $P(K)$ wird in den meisten naturwissenschaftlichen Gebieten mit 5 % angegeben, d.h., $P(K) \leq 0,05$.⁷⁴

Da lediglich die Entscheidung für G aus statistischer Sicht „ziemlich sicher“ ist⁷⁵, versucht man, den Fehler zweiter Art zu vermeiden. Dies ist nur gewährleistet, wenn man sich bei der Auswertung der Ergebnisse für die Gegenhypothese entscheidet. Somit sollte die eigentliche Hypothese als Gegenhypothese ausgesprochen werden. Man nimmt also an, dass die beobachtete zufällige Größe dem Modell der Hypothese H entspricht – liegt dann der beobachtete Wert im Bereich der kritischen Wahrscheinlichkeit $P(K)$, sieht man dies als statistischen Beweis dafür an, dass die Hypothese H ungültig, die Gegenhypothese G dagegen gültig ist.

⁷⁴ Stahel 1995: 175.

⁷⁵ Autorenkollektiv k.A.: 421.

5.2 Die Binomialverteilung in Bezug auf den ABX-Hörvergleich

Wie bereits zuvor ausführlich beschrieben, wird beim vorliegenden ABX-Hörvergleich getestet, ob ein Proband ein Signal X, welches mittels eines Zufallsgenerators aus den Quellen A und B bei jedem durchgeführten Experiment erneut bestimmt wird, der entsprechend richtigen Quelle A oder B zuordnen kann (also: entweder $X=A$ oder $X=B$). Mit X (hier nicht das Signal X!) wird die Anzahl der korrekten Entscheidungen des Probanden gezählt. Als Ereignis wird hier also das Auftreten einer korrekten Zuordnung gewertet.

Ist für den Probanden kein Unterschied hörbar, so entscheidet er sich willkürlich für eine der beiden Quellen A und B. Er entscheidet sich demzufolge mit einer Wahrscheinlichkeit von $p = q = \frac{1}{2}$ für die richtige Quelle (oder die falsche).

Entschließt sich der Testhörer bei sämtlichen Durchführungen des Experiments rein zufällig, so lässt sich hier eine Wahrscheinlichkeitsverteilung für $p = q = \frac{1}{2}$ angeben durch:

$$P(X=x) = p_x = \binom{n}{x} * \left(\frac{1}{2}\right)^n \quad \text{mit } x = 0, 1, \dots, n$$

Wegen der oft propagierten und häufig beschriebenen klanglichen Unterschiede zwischen DSD und High-Resolution-PCM wurde diese Annahme übernommen, aber – bezogen auf diesen Hörvergleich – aufgrund der oben erwähnten Vorgehensweise aus statistischen Gründen die Hypothese formuliert, dass die Probanden zwischen den Quellen A und B keinen Unterschied wahrnehmen, sich also rein willkürlich oder eben zufällig entscheiden. Demzufolge wurden folgende Aussagen aufgestellt:

- Hypothese H: Es existieren keine wahrnehmbaren Unterschiede zwischen der Quelle A und der Quelle B
- Gegenhypothese G: Es existieren wahrnehmbare Unterschiede zwischen A und B

5 Statistische Auswertungsmethode

Da sowohl von der International Telecommunication Union (ITU)⁷⁶ als auch von der ABX-Company⁷⁷ ein Signifikanz-Niveau von 5% empfohlen wird, lässt sich die kritische Wahrscheinlichkeit $P(K)$ mit $p = q = \frac{1}{2}$ nach folgender Gleichung bestimmen:

$$P(K) = \sum_{k=c}^n P(X = c) = \left(\frac{1}{2}\right)^n * \sum_{k=c}^n \binom{n}{k} \leq 0,05 \quad \text{mit } k = 0, 1, \dots, n$$

c: Anzahl der korrekten Entscheidungen

n: Anzahl der durchgeführten Versuche

$\sum_{k=c}^n P(X = c)$: Eintrittswahrscheinlichkeit $P(X \geq c)$, sie gibt an, wie hoch die Wahrscheinlichkeit ist, in n Versuchen mindestens c korrekte Antworten zu geben

Die ABX-Company empfiehlt eine Mindestanzahl von 16 durchzuführenden Versuchen. Trotz der größeren Anstrengung für den Probanden wurde die Zahl der Versuche auf 20 erhöht, da die Aussagekraft der statistischen Auswertung natürlich umso höher ist, je mehr Versuche durchgeführt werden (s. Tabelle der Eintrittswahrscheinlichkeiten am Ende dieses Kapitels).

Anhand von zwei Beispielen soll nun die Auswertung und Interpretation von Testergebnissen verdeutlicht werden:

Fall 1: Berechnung der Eintrittswahrscheinlichkeit für mindestens 14 korrekte Entscheidungen bei 20 durchgeführten Versuchen:

$$P(X \geq 14) = \left(\frac{1}{2}\right)^{20} * \sum_{k=14}^{20} \binom{20}{k} = 0,058$$

Bei rein willkürlicher Entscheidung für A oder B beträgt die Wahrscheinlichkeit für mindestens 14 korrekte Entscheidungen bei 20 Versuchen also 5,8%.

$P(X \geq 14)$ liegt also gerade eben nicht im Bereich der kritischen Wahrscheinlichkeit $P(K) \leq 0,05$. In diesem Fall würde man die Hypothese H beibehalten und die Gegenhypothese G verwerfen. Man könnte also die Vermutung äußern,

⁷⁶ ITU-R BS.1116-1 1997: 27.

⁷⁷ <http://www.pcavtech.com/abx/>.

5 Statistische Auswertungsmethode

dass ein Unterschied zwischen den Quellen A und B vom Probanden nicht wahrgenommen wird. Es sei aber an dieser Stelle noch einmal erwähnt, dass hier ein Fehler zweiter Art (Entscheidung für H, obwohl G zutrifft) nicht auszuschließen ist.

Fall 2: Berechnung der Eintrittswahrscheinlichkeit für mindestens 15 korrekte Entscheidungen bei 20 durchgeführten Versuchen:

$$P(X \geq 15) = \left(\frac{1}{2}\right)^{20} * \sum_{k=15}^{20} \binom{20}{k} = 0,021$$

Bei rein willkürlicher Entscheidung für A oder B beträgt die Wahrscheinlichkeit für mindestens 15 korrekte Entscheidungen bei 20 Versuchen also 2,1%.

Damit liegt $P(X \geq 15)$ gerade im Bereich der kritischen Wahrscheinlichkeit $P(K) \leq 0,05$. Es ergibt sich, dass man aufgrund der Entscheidungsregel die Hypothese H verwerfen und die Gegenhypothese G annehmen würde. In diesem Fall könnte man also davon ausgehen, dass der Proband einen Unterschied zwischen den Quellen A und B wahrgenommen hat.

Sicherlich mag es ein wenig befremdlich erscheinen, dass die Differenz von nur einer korrekten Antwort bei der Auswertung eine Rolle spielen kann, aber ein Schwellwert muss festgelegt werden – dieser liegt durch Konvention bei 5%.

Aus der nachfolgenden Grafik-Übersicht Nr.41⁷⁸ können – neben anderen – die verschiedenen Eintrittswahrscheinlichkeiten bei 20 durchgeführten Versuchen entnommen werden:

⁷⁸ Die enthaltenen Werte wurden von http://www.pcavtech.com/abx/abx_bino.htm übernommen.

5 Statistische Auswertungsmethode

Anzahl der Versuche	Anzahl der Versuche mit korrekter Entscheidung																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0,500																			
2	0,750	0,250																		
3	0,875	0,500	0,125																	
4	0,938	0,688	0,312	0,063																
5	0,969	0,813	0,500	0,188	0,031															
6	0,984	0,891	0,656	0,344	0,109	0,016														
7	0,992	0,938	0,773	0,500	0,227	0,062	0,008													
8	0,996	0,965	0,855	0,637	0,363	0,145	0,035	0,004												
9	0,998	0,980	0,910	0,746	0,500	0,254	0,090	0,020	0,002											
10	0,999	0,989	0,945	0,828	0,623	0,377	0,172	0,055	0,011	0,001										
11	1,000	0,994	0,967	0,887	0,726	0,500	0,274	0,113	0,033	0,006	0,000									
12	1,000	0,997	0,981	0,927	0,806	0,613	0,387	0,194	0,073	0,019	0,003	0,000								
13	1,000	0,998	0,989	0,954	0,867	0,709	0,500	0,291	0,133	0,046	0,011	0,002	0,000							
14	1,000	0,999	0,994	0,971	0,910	0,788	0,605	0,395	0,212	0,090	0,029	0,006	0,001	0,000						
15	1,000	1,000	0,996	0,982	0,941	0,849	0,696	0,500	0,304	0,151	0,059	0,018	0,004	0,000	0,000					
16	1,000	1,000	0,998	0,989	0,962	0,895	0,773	0,598	0,402	0,227	0,105	0,038	0,011	0,002	0,000	0,000				
17	1,000	1,000	0,999	0,994	0,975	0,928	0,834	0,685	0,500	0,315	0,166	0,072	0,025	0,006	0,001	0,000	0,000			
18	1,000	1,000	0,999	0,996	0,985	0,952	0,881	0,760	0,593	0,407	0,240	0,119	0,048	0,015	0,004	0,001	0,000	0,000		
19	1,000	1,000	1,000	0,998	0,990	0,968	0,916	0,820	0,676	0,500	0,324	0,180	0,084	0,032	0,010	0,002	0,000	0,000	0,000	
20	1,000	1,000	1,000	0,999	0,994	0,979	0,942	0,868	0,748	0,588	0,412	0,252	0,132	0,058	0,021	0,006	0,001	0,000	0,000	0,000

Grafik Nr.41: Tabelle der Eintrittswahrscheinlichkeiten. Hervorgehoben sind die Eintrittswahrscheinlichkeiten, die im Bereich des Signifikanzniveaus von 5% liegen. Die gerahmte Zeile gilt für den vorliegenden Fall bei 20 durchgeführten Versuchen.

6 Auswertung der Versuchsergebnisse

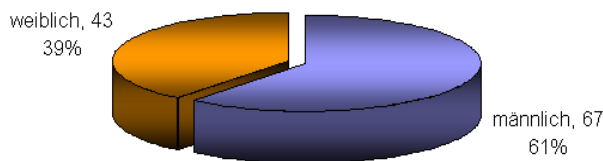
Ziel des vorliegenden Hörvergleichs ist, herauszufinden, ob es Probanden nachweisbar möglich ist, zwischen den zwei digitalen Kodierungsverfahren DSD und PCM (176,4kHz/24Bit) zu differenzieren. Dies sollte anhand von acht Stereo- und zwölf Surround-Musikbeispielen, die eigens für diesen Hörvergleich aufgenommen wurden, mit Hilfe eines ABX-Doppelblindtests untersucht werden. Die mathematische Auswertung basiert auf dem stochastischen Modell der Binomialverteilung.

Neben einer rein deskriptiven Auswertung soll in Kapitel 6.2 ergänzend eine interpretative Auswertung auf der Grundlage wahrnehmungspsychologischer Überlegungen erfolgen – es sei jedoch an dieser Stelle bereits deutlich herausgestellt, dass der Schwerpunkt der Auswertung eindeutig auf dem Kapitel 6.1 liegt.

6.1 Deskriptive Auswertung

Die Hörtests fanden innerhalb von 28 Testtagen statt. In diesem Zeitraum konnten 145 Tests mit 110 Testpersonen⁷⁹ durchgeführt werden. Dadurch wird auf jeden Fall den ITU-Richtlinien entsprochen, die besagen, dass bereits aufgrund der Ergebnisse von 20 Personen Schlussfolgerungen gezogen werden können.⁸⁰

Das Testpublikum bestand aus 43 weiblichen und 67 männlichen Testpersonen, deren



Altersstruktur aus Grafik Nr.43 ersichtlich wird. Im arithmetischen Mittel ergibt sich daraus ein Durchschnittsalter der Probanden von 32,9 Jahren.

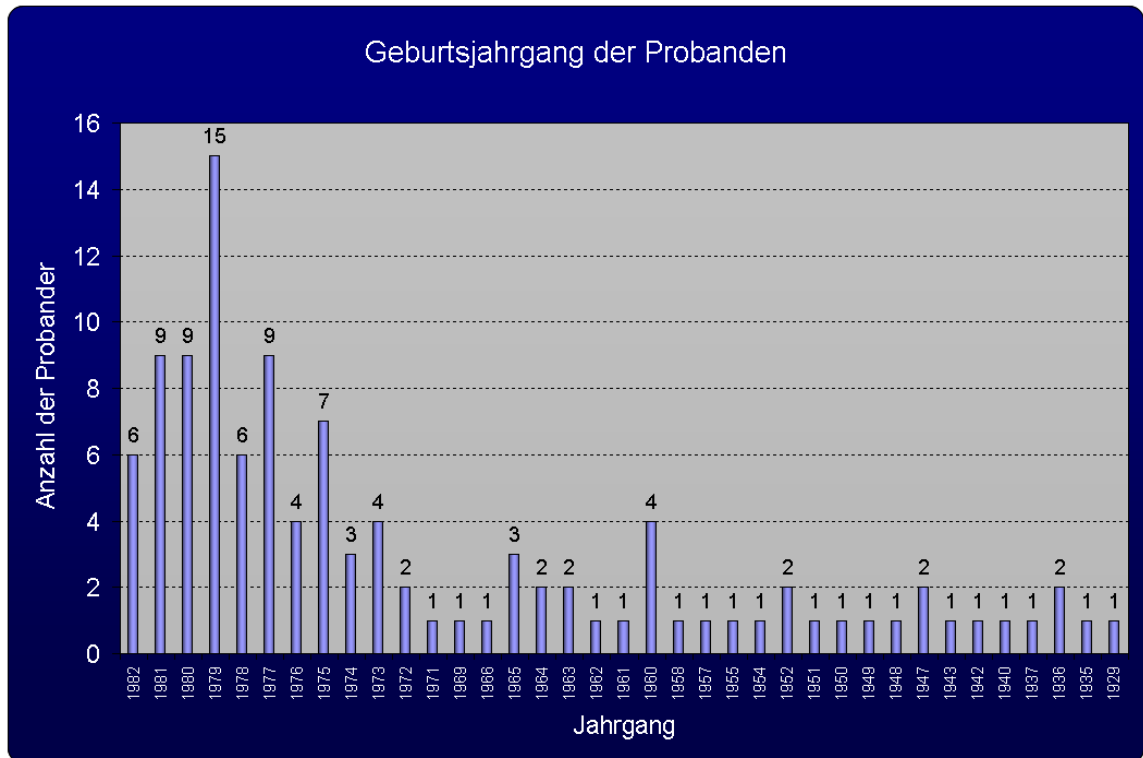
Grafik Nr.42:

Prozentuale Zusammensetzung des Testpublikums nach deren Geschlecht.

⁷⁹ Manche Probanden führten den Test mit zwei Musikbeispielen entweder direkt aneinander anschließend oder auch an unterschiedlichen Tagen durch.

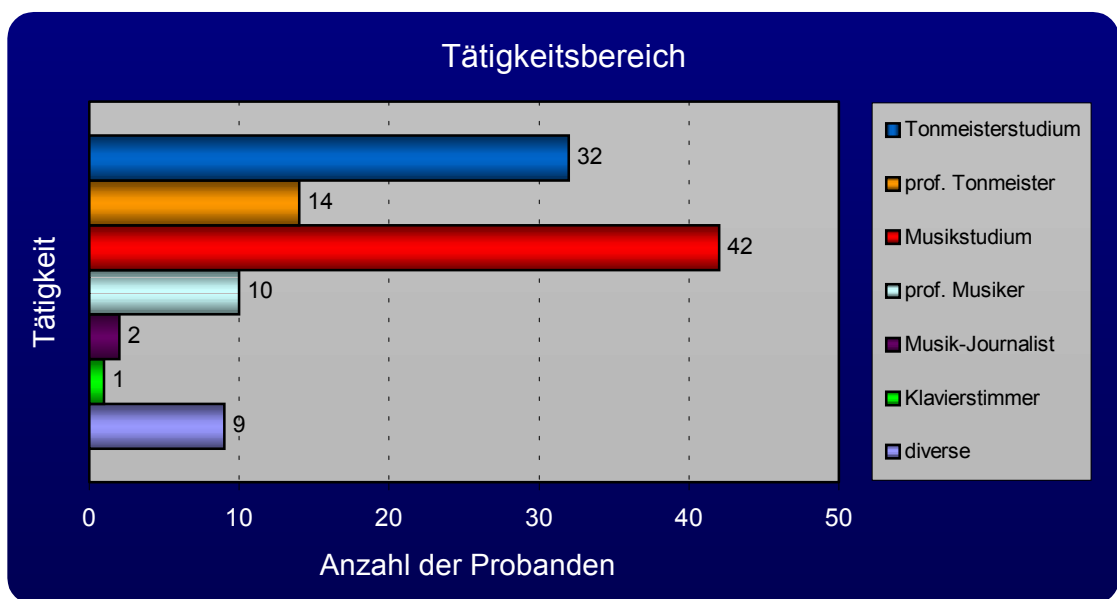
⁸⁰ ITU-R BS.1116-1 1997: 5.

6 Auswertung der Versuchsergebnisse



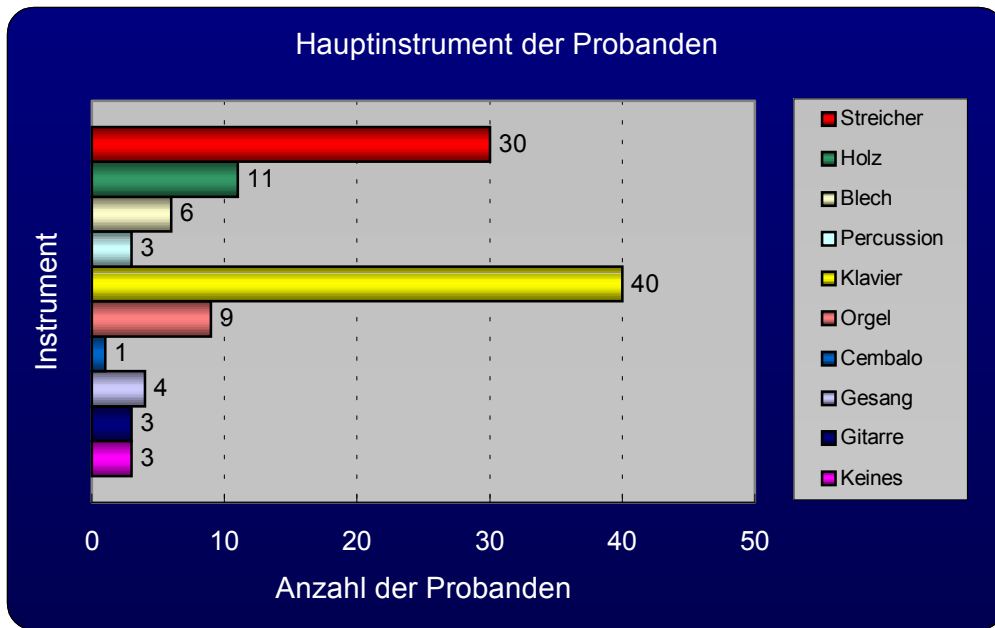
Grafik Nr.43: Altersstruktur der Testpersonen.

Grafik Nr.44 verdeutlicht, aus welchem beruflichen Betätigungsfeld sich die Testhörerschaft zusammensetzte. Ergänzend dazu bietet Grafik Nr.45 eine Übersicht über das jeweilige Hauptinstrument der Testpersonen. Aus beiden Grafiken lässt sich erkennen, dass das Testpublikum in der Mehrzahl auf jeden Fall (professionell-) musikalisches und kritisch-analytisches Hören gewohnt war.



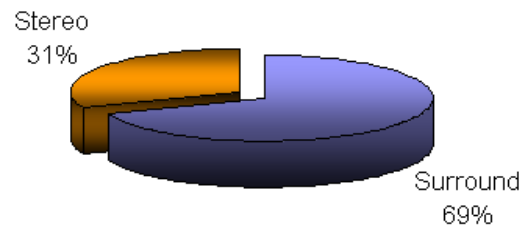
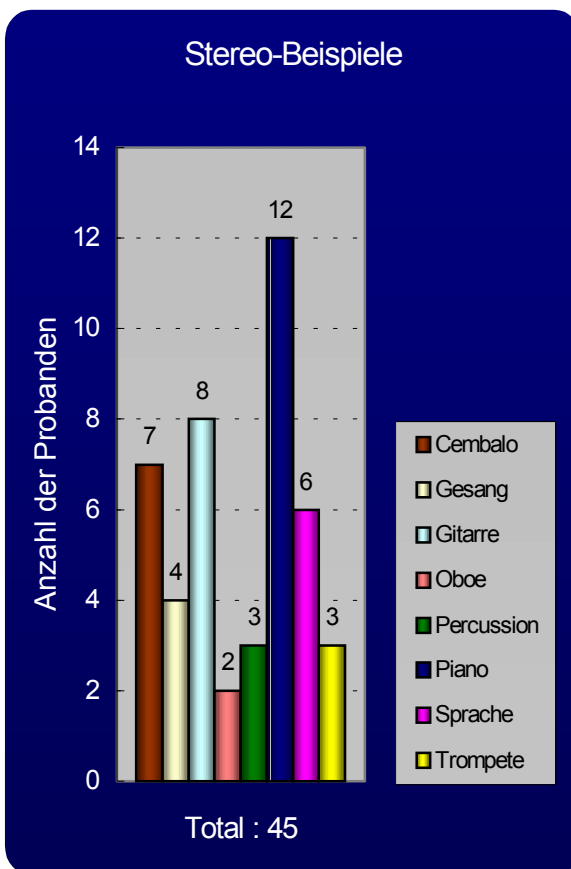
Grafik Nr.44: Beruflicher Tätigkeitsbereich der Probanden.

6 Auswertung der Versuchsergebnisse



Grafik Nr.45: Hauptinstrument der Probanden.

Die 145 durchgeführten Tests teilen sich in einem Verhältnis von 1:2,2 auf 45 Stereo-Beispiele (davon 30 mit Kopfhörerwiedergabe) und 100 Surround-Beispiele auf.

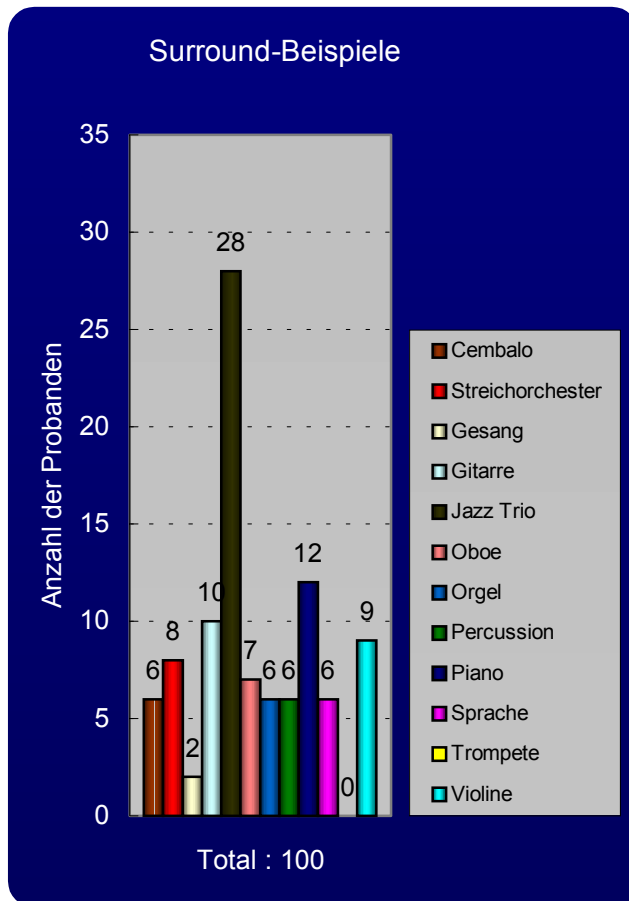


Grafik Nr.46: Verhältnis von Stereo- und Surround-Beispielen.

Aus den Grafiken Nr.47a und 47b lässt sich entnehmen, mit welchem Verteilungsverhältnis sich die Probanden für die 20 zur Auswahl stehenden Musikbeispiele in der eigentlichen Testphase entschieden haben.

Grafik Nr.47a: Verteilung der insgesamt 45 durchgeführten Stereo-Tests auf die zur Verfügung stehenden Musikbeispiele.

6 Auswertung der Versuchsergebnisse

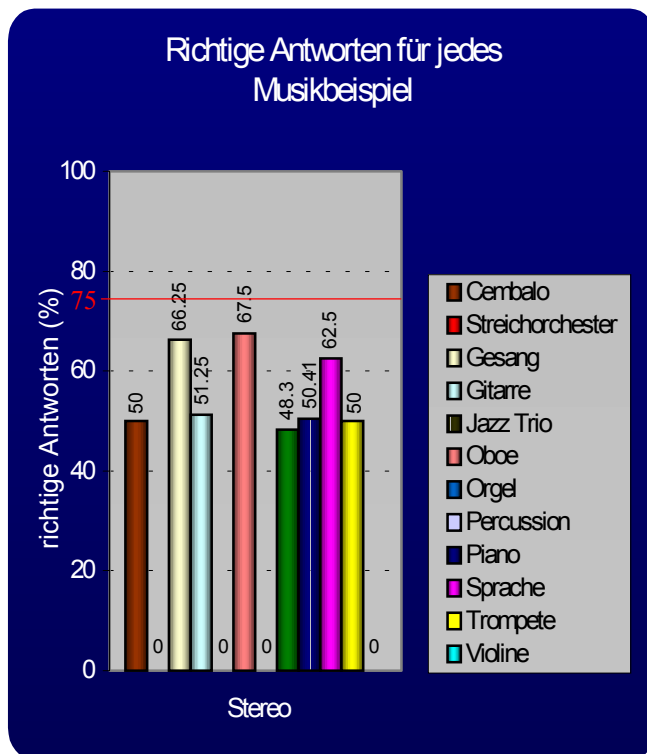


Grafik Nr.47b:

Verteilung der insgesamt 100 durchgeführten Surround-Tests auf die zur Verfügung stehenden Musikbeispiele.

Auffällig ist, dass das Jazz Trio überdurchschnittlich oft als Testbeispiel gewählt wurde. Begründet wurde das von den Probanden damit, dass durch die deutliche Panorama-Verteilung der Instrumente (Klavier ↔ L, R; Bass ↔ C; Schlagzeug ↔ LS, RS) die Aufnahme sehr durchsichtig sei und gleichzeitig sowohl unterschiedliche klangliche als auch räumliche Aspekte beinhalte und somit gute Anhaltspunkte zum Hören biete.

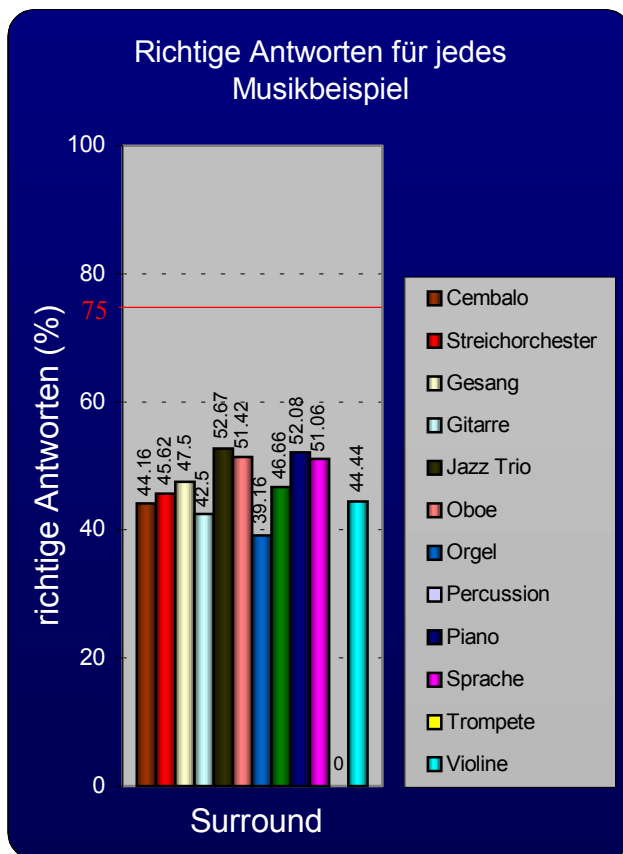
Wie sich die Anzahl der richtigen Antworten pro durchgeführtem Test im arithmetischen Mittel auf die einzelnen, verfügbaren Musikbeispiele in Stereo und Surround verteilt, ist aus den Grafiken Nr.48a und 48b ersichtlich.



Die horizontal rot eingetragene Linie markiert bei 75% (mit 20 Entscheidungen sind nur jeweils 5%-Schritte möglich) die untere Grenze der mindestens 15 richtigen Entscheidungen, bei denen die Eintrittswahrscheinlichkeit $P(X \geq 15)$ im Bereich der kritischen Wahrscheinlichkeit von 5% liegt.

Grafik Nr.48a: Anzahl der richtigen Antworten im arithmetischen Mittel für jedes Stereo-Musikbeispiel.

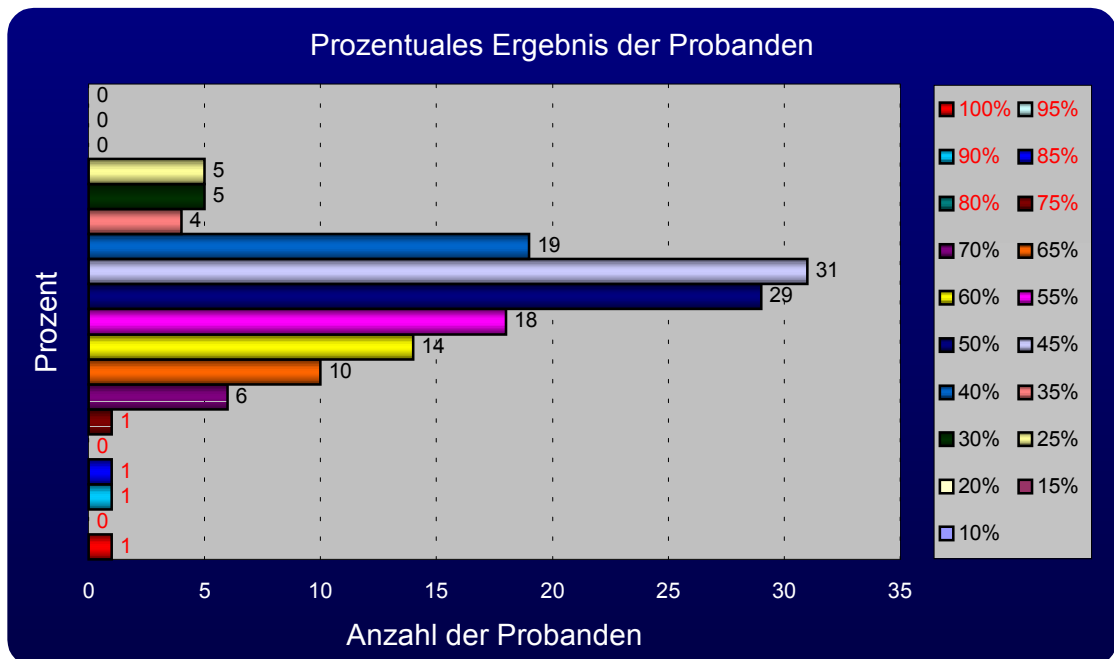
6 Auswertung der Versuchsergebnisse



Grafik Nr.48b:

Anzahl der richtigen Antworten im arithmetischen Mittel für jedes Surround-Musikbeispiel.

Aus Grafik Nr.49 lässt sich die Verteilung der 145 prozentualen Ergebnisse der Probanden entnehmen, die im Verlaufe der Testdurchführung erzielt wurden.



Grafik Nr.49: Anzahl der Probanden in Abhängigkeit von deren erreichtem prozentualen Ergebnis (Anzahl der richtigen Entscheidungen).

6 Auswertung der Versuchsergebnisse

Wie in Kapitel 5 ausführlich erklärt wurde, muss der Proband bei 20 Entscheidungen mindestens 75% (also 15 von 20) korrekte Antworten gegeben haben, um die kritische Wahrscheinlichkeit zu erreichen. Die Wahrscheinlichkeit hierfür beträgt $p = 0,021$.

Aus der Übersichtsgrafik Nr.49 wird deutlich, dass 4 Testpersonen im Bereich der kritischen Wahrscheinlichkeit liegen (rot unterlegt). Dies entspricht lediglich 2,76% der Gesamttests. Gemeinsam ist diesen 4 Ergebnissen, dass sie mit Stereo-Beispielen stattgefunden haben, die über Kopfhörer, also unter weitestgehender Ausblendung von ablenkenden Umwelt- und Raumeinflüssen, wiedergegeben wurden. Jeder dieser Probanden hatte dabei allerdings ein anderes Beispiel gewählt:

- **Oboe:** Stereo Kopfhörer **75%** richtige Antworten $\rightarrow p = 0,021$
- **Sprache:** Stereo Kopfhörer **85%** richtige Antworten $\rightarrow p < 0,005$
- **Gitarre:** Stereo Kopfhörer **90%** richtige Antworten $\rightarrow p < 0,005$
- **Gesang:** Stereo Kopfhörer **100%** richtige Antworten $\rightarrow p < 0,000...$

Leider wurden alle 4 Tests innerhalb der letzten vier Testtage absolviert, so dass es aufgrund einer hundertprozentigen Frequentierung der Testtermine in dieser Zeit nicht mehr möglich war, mit diesen Personen jeweils einen nachfolgenden Verifikationstest durchzuführen. Dies wäre mit Sicherheit interessant gewesen, jedoch nach Meinung der Testleiter keine zwingend notwendige Maßnahme, da die Aussagekraft der statistischen Auswertung aufgrund der – gegenüber den Empfehlungen – schon erhöhten Anzahl der zu treffenden Entscheidungen bereits größer als gefordert ist und somit reine „Zufallsergebnisse“ weitestgehend ausgeschlossen werden können (Zufallsergebnisse sind natürlich nicht ganz auszuschließen, die Wahrscheinlichkeit für diese ist jedoch äußerst gering). Für diese 4 Fälle soll im kommenden Unterkapitel 6.2.2 ein möglicher Erklärungsansatz skizziert werden, der aber bei der rein deskriptiven Auswertung keine Berücksichtigung findet.

Die Schlussfolgerung muss also lauten, dass bei diesen 4 Fällen im Bereich der kritischen Wahrscheinlichkeit aufgrund der Entscheidungsregel die Hypothese H (es existieren keine wahrnehmbaren Unterschiede zwischen der Quelle A und der Quelle B) verworfen und die Gegenhypothese G (es existieren wahrnehmbare Unterschiede zwischen A und B) angenommen werden kann. In diesen Fällen könnte davon ausgegangen werden, dass die Probanden einen Unterschied zwischen den Quellen A und B wahrgenommen haben.

6 Auswertung der Versuchsergebnisse

Dahingegen kann bei 141 von 145 Testergebnissen (entspricht 97,24%) die Hypothese H nicht abgelehnt, sondern muss beibehalten werden. In diesen Fällen könnte die Vermutung geäußert werden, dass ein Unterschied zwischen den Quellen A und B vom Probanden nicht wahrgenommen wurde.

Die Grafiken Nr. 48a und 48b zeigen sehr anschaulich, wie sich die prozentuale Verteilung bei den einzelnen Musikbeispielen in der Nähe des Zufalls-Niveaus von 50 Prozent bewegt (bei den Surround-Beispielen noch erheblich deutlicher als bei den Stereo-Beispielen). Bestärkt werden diese Beobachtungen, wenn alle falschen und richtigen Entscheidungen zusammen addiert werden: So ergibt sich bei insgesamt 145 Tests (multipliziert mit jeweils 20 Entscheidungen) eine Gesamt-Entscheidungszahl von 2900. Mit Hilfe der Grafik Nr.49 lässt sich eine Verteilung von 1454 richtigen zu 1446 falschen Entscheidungen berechnen, wobei eine bemerkenswerte Genauigkeit bezüglich des statistischen Erwartungswertes⁸¹ von 1450⁸² richtigen bzw. falschen Entscheidungen erreicht wird – bei nur 4 (nicht zu verwechseln mit den 4 signifikanten Testergebnissen) von diesem Erwartungswert abweichenden „Differenzentscheidungen“ entspricht das einer Abweichung von 0,28%.

Wie im Kapitel 4.4 erwähnt, wurde nach den ersten 10 getroffenen Zuweisungen von „X“ ein Zwischenergebnis abgespeichert, um eine eventuelle Entwicklung innerhalb des Testdurchlaufs festzustellen. Hier ließ sich jedoch keinerlei Regelmäßigkeit entdecken.

Da eine Differenzierbarkeit nur bei Stereo-Beispielen erfolgt ist, kann resümierend gesagt werden, dass Unterschiede bei den Surround-Beispielen, also unter Einbeziehung räumlicher Komponenten, in keinem von 100 Tests nachweislich wahrgenommen wurden. Ebenso konnte eine Unterscheidung perkussiver Elemente (Percussion und Cembalo), die sehr schnelle Signalanstiegsflanken enthalten, nicht vorgenommen werden. Es muss zusammenfassend festgestellt werden, dass bei dem vorliegenden Hörvergleich in der Regel nicht nachzuweisen war, dass klangliche Unterschiede zwischen DSD und High-Resolution-PCM (176,4kHz/24Bit) gehört wurden.

⁸¹ Arithmetischer Mittelwert einer Zufallsvariablen.

⁸² Bei Betrachtung des Tests als Zufallsexperiment.

6.2 Interpretative Auswertung

Über die rein deskriptive Auswertung hinaus soll in diesem Kapitel neben einigen allgemeinen Bemerkungen zur Diskussion um hochauflösende Aufzeichnungs- oder Kodierungsverfahren vor allem der Versuch einer möglichen Erklärung der vier – außerhalb des deutlich erkennbaren Trends liegenden – Testergebnisse unternommen werden⁸³. Es sei jedoch ausdrücklich betont, dass die Gültigkeit der Ergebnisse dieser vier Testpersonen keinesfalls angezweifelt wird; im Folgenden soll lediglich mit dem Wissen um die beim Testsetup entstandenen Schwierigkeiten ein möglicher Erklärungsansatz geliefert werden.

6.2.1 Allgemeine Vorbemerkungen

Immer wieder steht die Hörbarkeit von Ultraschall-Komponenten oberhalb des Audiobandes im Mittelpunkt von Diskussionen, in denen die Existenzberechtigung hoher Quantisierungs- und Samplingraten in Bezug auf die zwangsläufig entstehenden enormen (unkomprimierten) Datenmengen in Frage gestellt bzw. gerechtfertigt wird.

Einerseits gilt eine obere menschliche Wahrnehmungsgrenze von durchschnittlich 16-17kHz als allgemein akzeptiert, wobei sich diese Werte rein auf periodische Stimuli beziehen und Einzelpersonen bis ca. 20kHz oder auch darüber wahrnehmen können. Andererseits sind Schallereignisse mit unperiodischer Struktur für das menschliche Hörvermögen selbstverständlich im musikalischen Alltag weitaus verbreiteter, durch deren komplexe Strukturen jedoch ungeeigneter zur Feststellung einer oberen Wahrnehmungsgrenze. Des Weiteren können impulshafte Signale durch die steilen Anstiegsflanken in ihrem Einschwingvorgang im Zeitbereich Frequenzen oberhalb des Audiobandes entsprechen.

Daraus folgt die Vermutung, dass zwar bezogen auf periodische Signale eine individuell verschiedene Hörschwelle zwischen ca. 16 und 20kHz besteht, während des Einschwingvorgangs sehr steiler Impulse mit spektralen Anteilen oberhalb dieser Hörschwelle die Stereocilien⁸⁴ allerdings so stimuliert werden, dass diese Frequenzen

⁸³ Zumal bei 100 durchgeführten Surround-Tests kein Ergebnis im Bereich des Signifikanzniveaus lag.

⁸⁴ Beim Eintreffen eines Schallereignisses werden in der Cochlea im Innenohr jene auf der Basilarmembran befindlichen Stereocilien (Haarzellen) auf den Rezeptoren stimuliert, die auf die Weiterleitung über afferente Nervenfasern einer definierten Frequenz ausgerichtet sind.

6 Auswertung der Versuchsergebnisse

detektiert und verarbeitet werden können – bezüglich dieser schnellen Einschwingvorgänge stehen aber noch immer viele Fragen offen.

Um sich der menschlichen Perzeption bezüglich Frequenzen oberhalb des hörbaren Bereichs zu nähern, wurden bereits einige Hörversuche durchgeführt. Bedingt durch die Nichtlinearität von Lautsprechersystemen können sich bei der Wiedergabe über einen Lautsprecher jedoch Intermodulationsprodukte im wahrnehmbaren Bereich bilden, so dass sich ein hörbarer Unterschied rein auf technische Unzulänglichkeiten des Wiedergabesystems zurückführen ließe.

Dies bestätigen Untersuchungen, die im Rahmen der 110th AES-Convention in Amsterdam von Ashihara Kaoru und Kiryu Shogo präsentiert wurden.⁸⁵ Die beiden Autoren stellten fest, dass alle Probanden nur dann in der Lage waren, zwischen einem gegebenen Stimulus⁸⁶ mit und ohne Frequenzkomponenten oberhalb von 22kHz zu differenzieren, wenn dieser über einen einzelnen Lautsprecher wiedergegeben wurde. Sobald der Stimulus – aufgeteilt in 6 Frequenzbänder – zur Reduktion von Intermodulations-Verzerrungen über sechs, den einzelnen Frequenzbändern zugeordnete Lautsprecher wiedergegeben wurde, war es den Probanden nicht mehr möglich, Ultraschallkomponenten zu detektieren. Daraus zogen sie die Schlussfolgerung, dass das Hinzufügen von Ultraschallkomponenten eventuell Klangeindrücke hervorruft, die durch nichtlineare Produkte in den Lautsprechern verursacht werden.

In einer auf der 115th AES-Convention in New York vorgestellten Studie der NHK Science & Technical Research Laboratories, Tokyo⁸⁷ wurde mit Hilfe von Hörversuchen die Unterscheidbarkeit von Musik mit und ohne Frequenzanteilen oberhalb des Audiobandes geprüft. Um Artefakte durch Intermodulations-Verzerrungen zu unterbinden, wurden zwei vollkommen unabhängige Wiedergabepfade verwendet: Zum einen für das hörbare Audioband, zum anderen für die Ultraschallfrequenzen. Jedes der 20 für den Test angebotenen Musikbeispiele wurde mittels eines digitalen FIR-Tief- bzw. Hochpassfilters mit sehr steiler Roll-Off-Charakteristik in zwei Frequenzbänder unterteilt. Während der Hörversuche wurde das hörbare Audioband permanent, das Ultraschallband nur temporär wiedergegeben. Die Ergebnisse zeigen, dass von 36

⁸⁵ AES Convention Paper #5401 2001.

⁸⁶ „A harmonic complex tone consisted of only odd number harmonics was used as the stimulus. Its fundamental frequency (f_0) was 2kHz. The lower 5 components that were the fundamental, 3rd, 5th, 7th and 9th harmonics were defined as non-target components and the higher 5 components that were the 11th, 13th, 15th, 17th and 19th harmonics were defined as target-components. Frequency of the highest harmonics was 38kHz. The duration of the stimulus was 2000ms including linear onset and offset ramps of 100ms each.“ [AES Convention Paper #5401 2001: 1].

⁸⁷ AES Convention Paper #5876 2003.

6 Auswertung der Versuchsergebnisse

Testpersonen lediglich eine in der Nähe der kritischen Wahrscheinlichkeit lag. Bei einer Wiederholung mit 6 unterschiedlichen Beispielen war es dieser Probandin jedoch nicht möglich, eine Differenzierung zwischen den einzelnen Signalen (mit und ohne Ultraschallanteilen) im Bereich des Signifikanzniveaus vorzunehmen. Hieraus schlossen die Autoren, dass die Möglichkeit der Differenzierbarkeit beider Signale durch Probanden weder bestätigt noch verneint werden kann.

Darüber hinaus sei aus hörphysiologischer Sicht angemerkt, dass unser hauptsächlich zur Verstärkung der eintreffenden Reize dienendes Mittelohr, bestehend aus den Gehörknöchelchen Hammer, Amboss und Steigbügel, einen sehr steilen mechanischen Tiefpass darstellt. Dadurch kann nur sehr wenig Ultraschallenergie vom Trommelfell überhaupt zum Innenohr vordringen, somit auch nur in einem geringen Maße von der Cochlea aufgenommen und über die afferenten Nervenfasern weitergeleitet werden.

Des Öfteren wird bei Diskussionen das zeitliche Auflösungsvermögen eines digitalen Kodierungsformates, bedingt durch dessen Samplingrate, erwähnt und auf die Wahrnehmungsfähigkeit des menschlichen Gehörs im Bereich von nur wenigen Mikrosekunden verwiesen. Während eine Samplingrate von 44,1kHz alle 22,67 Mikrosekunden eine „Datenerneuerung“ durchführt, sinkt diese Zeitspanne bei einer Abtastrate von 192kHz bereits auf 5,2µsec – bei einer Abtastrate von 2,8224MHz liegt sie im Bereich von ca. 354nsec. Von der Tatsache abgesehen, dass die Phasendarstellung nicht durch das Zeitraster der Abtastfrequenz beeinflusst wird, sondern beliebig genau erfasst werden kann, entspricht die menschliche frequenzabhängige Vornelokalisationsunschärfe von zwei Grad bei Sinussignalen einem zeitlichen Auflösungsvermögen von ca. 20µsec.⁸⁸ Angemerkt sei ferner, dass bei der Überführung einer mechanischen Bewegung der Stereocilien in neuronale Aktivität ein elektrischer Impuls innerhalb von (nur) etwa 10µsec generiert wird.⁸⁹

Die vorliegenden Erkenntnisse der Forschung bezüglich der Wahrnehmbarkeit von Ultraschall sowie des zeitlichen Auflösungsvermögens sind jedoch teilweise unklar und werden äußerst kontrovers diskutiert.

⁸⁸ Zwicker, Fastl 1999: 309f.

⁸⁹ Spitzer 2002: 61.

6.2.2 Wahrnehmungspsychologische Überlegungen

Wie in den Ergebnissen der Auswertung des Hörvergleichs festgestellt wurde, war es insgesamt vier Probanden möglich, bei jeweils unterschiedlichen Musikbeispielen, jedoch übereinstimmend bei Kopfhörerwiedergabe (also Stereobeispielen), innerhalb der kritischen Wahrscheinlichkeit von 5% zu liegen. Im Folgenden soll mit dem Wissen um die technischen Probleme, die das umfangreiche Versuchsssetup mit sich gebracht hat, ein möglicher spekulativer Erklärungsansatz dieser vier – außerhalb des klar erkennbaren Trends liegenden – Fälle beschrieben werden:

Systembedingt produziert das Non-Audio-Format bei Funktionen, die direkt oder auch indirekt eine „Stop“- oder „Play“-Funktion beinhalten, beim DSD-Modus ein im Gegensatz zum PCM-Modus teilweise minimal unterschiedlich klingendes, sehr kurzes Knackgeräusch, das vermutlich nur durch einen Fade-In oder Fade-Out auf digitaler Ebene hätte vermieden werden können. Dies würde einerseits natürlich der Voraussetzung widersprechen, dass das Material in keiner Weise bearbeitet sein darf; der essentiellere Aspekt jedoch wäre andererseits, dass ein solcher nur schwer in einem Non-Audio-Format realisierbar ist. Trotz intensiver Beschäftigung mit diesem Problem und Beratung durch einige der den Hörvergleich unterstützenden Firmen, sowie einer Umstrukturierung des ursprünglich geplanten Rechnerkonzepts, war zum damaligen Zeitpunkt eine Lösung durch eine zusätzliche Soft- oder Hardware nicht greifbar und wäre auch momentan nur schwer umzusetzen.

Dass dieses Knacken nicht bewusst zur Kenntnis genommen wurde und der Test somit in keiner Weise seiner Aussagekraft enthoben wird, zeigt die Auswertung der Fragebögen, in denen ein Knackgeräusch ausnahmslos nicht erwähnt bzw. nicht eindeutig einem der beiden Signale zugewiesen wurde. Dies bestätigt die Annahme, dass durch die Anlage und die Zielsetzung des Hörvergleichs eine bewusste Hörfokussierung auf klangliche Parameter im Mittelpunkt stand, wohingegen es beiden Testleitern wiederholt möglich war, unabhängig vom gewählten Musikbeispiel durch eine bewusst gewählte auditive Aufmerksamkeitslenkung auf das Knackgeräusch eine eindeutige Zuordnung des Signals „X“ zum jeweiligen Quellsignal „A“ oder „B“ vorzunehmen. Aus diesem Grund soll im Folgenden näher auf Aufmerksamkeit und Aufmerksamkeitslenkung eingegangen werden.

Aufmerksamkeit ist ein „Zustand der gesteigerten Wachheit (Vigilanz) und Anspannung, der der selektiven Orientierung des Wahrnehmens, Denkens und Handelns zugrunde liegt. [...] Mit *Aufmerksamkeits-Regulation* bzw. *-Steuerung* [...] werden neurophysiologische Prozesse bezeichnet, welche die Aufnahme, Verarbeitung und Nutzung bestimmter Informationen fördern, der Auswahl (Selektion) der individuell bedeutsamen Situationselemente bei gleichzeitiger Dämpfung irrelevanter Eindrücke Nachdruck verleihen, [...]. Aufmerksamkeit gilt als Selektionsfunktion der Bewußtseinstätigkeit, die das Klarheitsrelief der Eindrücke in Analogie zum Wahrnehmungsfeld bestimmt. Im fokalen, d.h. zentralen Bereich erscheinen Einzelheiten klar und deutlich, so daß sie in ihrer gegenständlichen Bedeutung aufgefaßt werden können (*Apperception*); im peripheren Bereich dagegen gibt es bestenfalls mitbewußte, unklare Eindrücke (*Perceptionen*). Selektive Aufmerksamkeit bzw. Unaufmerksamkeit sind Voraussetzungen des geordneten Erkennens und Denkens, da Klarheit und Deutlichkeit notwendig sind und Aufmerksamkeit und Bewußtseinsumfang dem Erkennen deutliche Grenzen setzen.“⁹⁰

Zurückgehend auf den Ansatz von Broadbent wurde in den 1950er Jahren angenommen, dass Filter im Wahrnehmungssystem für die Verringerung der Informationsvielfalt zuständig sind, indem sie nicht alle Nervenreize weiterleiten. Diese Theorien werden als sogenannte Flaschenhals-Theorien (bottleneck theories) bezeichnet. Ausgangspunkt für diese ist die eingeschränkte Kapazität der zentralen Informationsverarbeitung.

Aufgrund der Ergebnisse von Untersuchungen zum dichotischen Hören⁹¹ formulierte Broadbent 1958 ein Filtermodell der selektiven auditiven Wahrnehmung: Seiner Meinung nach war die Aufmerksamkeit nach dem Prinzip eines selektiven Filters organisiert, das die Menge der eingehenden sensorischen Daten entweder durch Weitergabe wichtiger Informationen an das Bewusstsein, oder durch Sperrung unerwünschter Informationen steuert. Dabei gelangen die eintreffenden Daten vor dem eigentlichen Filter zunächst in eine Pufferzone. Werden Daten vom Filter blockiert, verbleiben sie im Pufferspeicher. Broadbent vermutete, dass das entscheidende Kriterium für das Passieren einer Mitteilung durch den Filter auf der Basis physischer Merkmale geschieht. Darüber hinaus nahm er an, dass nur Daten aus einem einzigen

⁹⁰ dtv-Wörterbuch zur Psychologie 1994: 75ff.

⁹¹ Die Organisation eines Experiments nach dem dichotischen Prinzip bedeutet die simultane Darbietung zweier unterschiedlicher Reize, wobei jedem Ohr jeweils ein Reiz dargeboten wird. Die Versuchspersonen sollen dabei einen Reiz von dem anderen unterscheiden. Die Fähigkeit zu dieser Leistung wird als "Beschatten" ("shadowing") bezeichnet.

6 Auswertung der Versuchsergebnisse

Kommunikationskanal⁹² gleichzeitig durchgelassen werden. Dieser Kanal ist in seiner Kapazität begrenzt.

Eine Modifikation der Filtertheorie ergab sich 1960 aufgrund von Beschattungsexperimenten durch Treisman, da manche Sachverhalte (wie z.B. der Cocktail-Party-Effekt) durch die Broadbent'sche Theorie nicht mit den Beobachtungen in Übereinstimmung zu bringen waren.

Treisman schaltete bei den Versuchspersonen eine auf dem linken Ohr zu beschattende Mitteilung plötzlich zum rechten, eigentlich zu ignorierenden Ohr um, während die auf dem linken Ohr zu hörende Mitteilung in eine zufällige, sinnlose Wortfolge wechselte. Dabei stellte er fest, dass die Versuchspersonen einige wenige Worte dieser Mitteilung zu reproduzieren in der Lage waren, obwohl diese dem nichtbeschatteten Ohr dargeboten wurden.

Basierend auf einem weiteren Beschattungsexperiment konnte Treisman 1964 seine nach dem ersten Experiment aufgestellte Annahme verifizieren, dass Mitteilungen nicht allein nach physischen, sondern auch nach semantischen Kriterien verarbeitet werden und formulierte die sogenannte Dämpfungstheorie der selektiven auditiven Aufmerksamkeit.

Die Kernaussage dieser Theorie ist, dass eingehende Daten auf Basis physischer Kriterien gedämpft, aber nicht völlig ausgefiltert werden; die semantischen Kriterien werden auf alle Informationen angewendet, ob beachtet oder nicht.

Ziel des vorliegenden Hörvergleichs ist, bei den zwei unterschiedlichen Kodierungsverfahren DSD und High-Resolution-PCM (176,4kHz/24Bit) in der klanglichen Wiedergabe Elemente der Differenzierung aufzufinden, so dass eine reproduzierbare Unterscheidung beider gegebenenfalls möglich ist. Dabei spielen klangliche Parameter wie z.B. Räumlichkeit, Stereo-Abbildung (Surround-Abbildung), Transparenz, Klangbalance, Klangfarbe, Impulsverhalten, etc. eine entscheidende Rolle.

Die Aufmerksamkeit richtet sich dabei also nicht auf semantische Aspekte der Musik (im Sinne einer Bedeutungsebene), sondern auf Aspekte des Klanges und seiner zugehörigen Elemente (u.a. auch Knacken, Rauschen, etc.).

⁹² Darunter versteht man z.B. Reizdaten mit bestimmten Eigenschaften, Daten aus einem bestimmten Sinnesgebiet bzw. -organ, oder Daten, die in einem Bedeutungszusammenhang stehen.

6 Auswertung der Versuchsergebnisse

In Analogie zu den sinnvollen Wortzusammenhängen im Fall des Treisman-Beschattungsexperiments lässt sich eine tonmeisterspezifische semantische Ebene – also eine Bedeutungsebene des Klanges – annehmen: Diese wird während der Ausbildung sowie der späteren Berufspraxis entwickelt und verfeinert. Ihre inhaltliche Bedeutung ist für die Aufmerksamkeitslenkung von TonmeisterInnen entscheidend.

Da alle vier im Bereich der kritischen Wahrscheinlichkeit liegenden Probanden Tonmeister(-studierende) waren und die jeweiligen Musikbeispiele bezeichnenderweise über Kopfhörer gehört und sich demzufolge noch einmal ausschließlicher auf das Hören eingelassen haben (störende Raum- bzw. Umwelteinflüsse werden deutlicher ausgeschlossen), könnte diese Bedeutungsebene hier besonders wirksam werden.

An dieser Stelle sei hinzugefügt, dass es mittels Gedächtnis-, Konditionierungs- und Reaktionszeituntersuchungen gelungen ist, einige Ergebnisse bezüglich der Verarbeitung der sekundären (der zu ignorierenden) Mitteilung zu erhalten. Auch wenn diese teilweise nicht eindeutig sind, lässt sich feststellen, dass die zu ignorierenden eingehenden Daten unterhalb der Bewusstseinschwelle auf bedeutungsrelevante Informationen hin überwacht werden und die selektive auditive Aufmerksamkeit somit auch von vorhandenen Wissensstrukturen des Individuums beeinflusst wird.⁹³

Betrachtet man das Testergebnis im Zusammenhang gerade mit der Dämpfungstheorie nach Treisman, die heute als weitestgehend anerkannt gilt, können auch klangliche Elemente wie ein Knacken als – tonmeisterspezifisch – semantische Inhalte gelten und eine Entscheidungsfindung unbewusst oder mitbewusst beeinflussen.

⁹³ Zimbardo 1995: 228.

7 Fazit

Aufgrund der Resultate des vorliegenden Hörvergleichs hat sich gezeigt, dass selbst mit hochwertigstem Equipment unter optimalen Abhörbedingungen und unterschiedlichster Hörfokussierungen bzw. Hörerfahrungen der Probanden in der Regel keine signifikanten Unterschiede zwischen DSD und High Resolution PCM (24bit/176,4kHz) hörbar sind, sich demzufolge die These aufstellen ließe, dass sich keines der getesteten Systeme durch klangliche Eigenschaften hervorhebt. Lediglich bei 4 der 145 durchgeführten Tests liegt das Ergebnis im Bereich des Signifikanzniveaus von 5%, übertragen auf die Gesamtauswertung also nur 2,76% der Testergebnisse. Während diese vier Tests sich ausschließlich auf Stereobeispiele bei Kopfhörerwiedergabe beziehen, lag bei 100 durchgeführten Surround-Tests kein Testergebnis im Bereich der kritischen Wahrscheinlichkeit. Ein möglicher Erklärungsansatz für diese deutlich außerhalb des klar zu erkennenden Trends liegenden vier Fälle wurde ausführlich in Kapitel 6.2 angedacht.

Wenn im Allgemeinen nicht auf auditiver Ebene eindeutige Differenzierungen vorgenommen werden können, liegt die Schlussfolgerung nahe, eine Abwägung nach rein technischen Gesichtspunkten vorzunehmen: Im Laufe des Siegeszuges der CD in den vergangenen zwei Jahrzehnten konnte sich die auf PCM basierende Digitaltechnik in sämtlichen Audibereichen durchsetzen und weiterentwickeln. Das betrifft digitale Mikrofone, Mischpulte, Effektgeräte, Lautstärke-Controller, etc. – diese Errungenschaften sind im gegenwärtigen Produktionsablauf kaum noch zu entbehren. Soll hingegen eine „reine“ DSD-Produktion stattfinden, muss aufgrund der fehlenden Kompatibilität des digitalen Ein-Bit-DSD-Datenstroms und des bis heute nur äußerst begrenzt zur Verfügung stehenden Equipments in der Regel auf jegliche Bearbeitungsmöglichkeiten verzichtet werden. Der Umweg über konventionelle Mischpulte, Effektgeräte und der Verzicht auf z.B. ein digitales Hallgerät ist momentan nicht zu vermeiden, ein steter Qualitätsverlust des DSD-Signals bei mehreren nacheinander ausgeführten Requantisierungsschritten ist die unweigerliche Folge.

Andererseits bietet die SACD als Trägerformat einen – zumindest für die Musikindustrie – entscheidenden Vorteil: Ein mehrfacher, aus einer Kombination physikalischer und kryptographischer Verfahren bestehender Kopierschutz bietet ein hohes Maß an Sicherheit vor dem Missbrauch unerlaubter Vervielfältigung, so dass die durch

die immer weiter steigende Zahl von Raubkopien finanziell stark in Mitleidenschaft gezogene Plattenindustrie eventuell eine Aussicht auf Rehabilitation hätte.

Ein möglicher Kompromiss, der sich aus diesem Hörvergleich und seiner Auswertung ergibt, wäre eine Kombination der Vorteile beider Systeme: So wäre es denkbar, den gesamten Aufnahme- und Postproduktionsprozess weiterhin mit dem bereits bestehenden PCM-Equipment durchzuführen, das fertige Audio-PCM-Material dann mittels eines „Digital-to-Digital“-Wandlers nach DSD zu konvertieren und, den effektiven Kopierschutz nutzend, auf einer SACD zu veröffentlichen – ein Ansatz, der sicherlich bereits realisiert wird und natürlich Vor- aber auch Nachteile birgt, die es abzuwägen gilt⁹⁴.

Die Auswertung des Fragebogens ergab, dass von den getesteten Personen (also interessiertem Fachpublikum) lediglich 7 Personen eine Surroundanlage besitzen. Betrachtet man jedoch das Verhältnis von gewählten Stereo- zu Surround-Beispielen, ist selbst unter Einbeziehung der Tatsache, dass mehr Surround- als Stereobeispiele angeboten wurden, noch eine klare Präferenz der Mehrkanal-Aufnahmen festzustellen. Dies positiv interpretierend, zeigt sich, dass ein grundlegendes Interesse seitens potentieller Konsumenten vorhanden ist und „Berührungängste“ nicht existieren. Es scheint daher sinnvoller, sich nicht nur auf diverse firmenpolitische und – vor dem Hintergrund der vorliegenden Ergebnisse – nicht bestätigte Aussagen zu den vermeintlich großen klanglichen Unterschieden zu beschränken. Vielmehr ist es im Interesse aller, v.a. auch vor dem Hintergrund einer bereits existierenden und stetig wachsenden Formatvielfalt, einen einheitlichen Standard zu finden und nicht die Durchsetzung von vorteilbringenden und zukunftssträchtigen Innovationen wie dem Surround-Sound durch Kundenverwirrung zu gefährden bzw. gleich im Keim zu ersticken.

Weniger nachweislich in Form einer mathematischen Auswertung, jedoch umso einprägender war das hohe Maß an Frustration, das viele Probanden, die in der Mehrzahl professionelles und kritisch-analytisches Hören gewohnt waren, während der Durchführung der Tests empfanden und das sie auf für sie nicht annähernd zu erkennende klangliche Unterschiede zurückführten. Ein weiteres, häufiges Diskussionsthema war ebenso die Wahrnehmung von *vermeintlichen* Unterschieden: Hörillusion – ein Thema,

⁹⁴ Wie sich ein finaler Konvertierungsprozess von PCM nach DSD klanglich auswirkt, oder ob dieser reproduzierbar wahrgenommen wird, könnte z.B. Gegenstand eines weiteren Hörvergleichs sein.

das sicherlich speziell für Tonmeister eine bedeutende Rolle spielen und über dessen Ursache – soll effektiv und sowohl musikalisch als auch klanglich gewinnbringend gearbeitet werden – persönliche Klarheit bestehen sollte. Gerade PR-Abteilungen wissen um die Relevanz von u.a. visuellen und informativen Assoziationen und den daraus resultierenden Erwartungshaltungen der Konsumenten. Den Aspekt der Zuverlässigkeit und Glaubwürdigkeit klanglich bezogener Aussagen betrachtend, stellte sich während der Tests sehr deutlich heraus, wie wichtig und sinnvoll Hörvergleiche in Form eines (Doppel-)Blindtests sind, da nur so festgestellt werden kann, was überhaupt noch reproduzierbar wahrgenommen wird.⁹⁵

Aus technischer Sicht bleibt sicherlich die Frage zu stellen, wie weit die beiden verglichenen Systeme wirklich noch voneinander entfernt sind – schließlich findet, zumindest was die empfohlene Bearbeitung des Audiosignals betrifft, durch DSD-Wide (bzw. „PCM-Narrow“⁹⁶) bereits eine Annäherung an das Multibit-PCM-Format statt – und es sollte diesbezüglich überlegt werden, wie gültig das einstige Postulat einer „fundamentalen Abkehr von Multibit-PCM“ noch ist.⁹⁷

Vor dem Hintergrund des bereits 1998 am Erich-Thienhaus-Institut durchgeführten Hörvergleichs, bei dem sich gezeigt hat, „daß zwischen hochqualitativen analogen und mit entsprechend guten 48 kHz/24 bit Wandlern digitalisierten Musiksignalen keine signifikanten Unterschiede hörbar sind“⁹⁸ und den Ergebnissen des vorliegenden Hörvergleichs sollte ebenfalls überlegt werden, wohin der derzeitige Trend einer zunehmenden Erhöhung der Samplingrate – jedenfalls aus klanglicher Sicht – führen soll, wenn dem daraus resultierenden Anwachsen der Datenmengen⁹⁹ andererseits Komprimierungsverfahren wie MP3, Dolby Digital AC3, WMA, dts, etc. gegenüberstehen. Inwieweit nicht eine stärkere Besinnung auf die eigentliche Tonmeisterarbeit – die Auswahl geeigneter Mikrofone, deren Positionierung, Schnitt, Mischung, eine musikalisch fundierte Produktion der aufzunehmenden Musiker, etc. – nicht erheblich größere klangliche Verbesserungen bewirken würde, als die Entscheidung für oder

⁹⁵ Dem interessierten Musikhörer ist es aufgrund des enormen technischen Aufwands zum gegenwärtigen Zeitpunkt leider nicht möglich, sich ein eigenes Hörurteil im Rahmen eines möglichst objektiven Vergleichstests zu bilden.

⁹⁶ St.Lipshitz in Müller 2002: 43.

⁹⁷ SACD – A Technical Proposal 1997: 6.

⁹⁸ Acker, Gericke 1998: 114.

⁹⁹ Wenngleich diese zukünftig durch die neue Formatvielfalt mit Blu-ray, AOD und UDO nicht das größte Problem darstellen werden.

7 Fazit

wider ein digitales Kodierungsverfahren (zumindest bei deren derzeitiger hoher Auflösung), bleibt zu diskutieren.

Auf jeden Fall sollte herausgestellt werden, dass sich aus Sicht der Tonmeister sowohl musikalisch als auch technisch für die Verbraucher sicherlich ein erheblicher Qualitätsgewinn dadurch ergibt, dass die während der Aufnahme gewonnene höhere Auflösung – unabhängig vom Distributionsformat – an den Verbraucher weitergegeben werden kann und es somit möglich ist, sich von den technischen Limitierungen der CD mit 44,1kHz und 16 Bit zu lösen.

Wir hoffen, mit der abgehandelten Untersuchung zur auditiven Differenzierbarkeit digitaler Aufzeichnungsverfahren einen förderlichen Beitrag zur aktuellen Diskussion um hochauflösende digitale Audioformate geleistet und Anstoß für weitere Überlegungen und nachfolgende Untersuchungen gegeben zu haben.

8 Danksagung

Diese Diplomarbeit war selbstverständlich nur durch verlässliche, gute Zusammenarbeit und v.a. Kooperationsbereitschaft vieler Personen möglich.

Da es uns aufgrund des großen Personenkreises leider nicht möglich ist, sämtliche Mitwirkenden im Einzelnen zu nennen, möchten wir uns ganz herzlich bei allen Beteiligten bedanken, die zu einer erfolgreichen Durchführung der Arbeit beigetragen haben, insbesondere jedoch bei

allen Firmen sowie deren Vertretern, nicht nur für viele interessante und informative Gespräche, sondern v.a. für die Bereitstellung des umfangreichen Equipments über einen sehr langen Zeitraum. Ohne diese in allen Fällen sehr spontane Hilfsbereitschaft hätte dieser Hörtest nicht stattfinden können. Hier sind – in der Reihenfolge des Signalflusses – zu nennen:

- Herr Wuttke und Herr Vollmer, **Schalltechnik Dr.-Ing. Schoeps GmbH** (Mikrofone)
- Herr Zielinsky, **Sennheiser electronic GmbH & Co. KG** (Mikrofone)
- Herr Reim, **Lake-People electronic GmbH** (Mikrofonvorverstärker)
- Herr van der Wolf, Herr Story und Herr Weber, **dCS Limited** (Konverter)
- Herr Ludwig, **Media Assistance** (Pyramix Soft- und Hardware)
- Herr Cellier, **Merging Technologies** (Pyramix Soft- und Hardware)
- Frau Schulz, Herr Kутtenkeuler und Herr Shibata, **Emil-Berliner-Studios** (Monitorunit, ABX-Testsoftware)
- Frau Manger und Herr Schuldt, **Manger Products** (Endstufen und Lautsprecher)
- Frau und Herr Zühlke, **Stax - Germany** (Kopfhörer mit -verstärker)
- Herr Klotz, **Klotz a.i.s. GmbH** (Kabel)
- Herr Holtz, **FH-Lippe**, Abteilung Elektrotechnik (Messungen)
- unser Kommilitone Christian Struck (ferromag-Netzfilter-Leiste)

den motivierten und geduldigen MusikerInnen, die sich für die zumeist nächtlichen Aufnahmen der Musikbeispiele bereitwillig zur Verfügung gestellt haben,

8 Danksagung

dem Hausdienst der Hochschule für Musik Detmold, der einmal wieder viel Verständnis und Hilfsbereitschaft gezeigt hat,

unseren Betreuern Rainer Maillard und Andreas Meyer für die vielen konstruktiven Überlegungen und Gespräche,

unserem Studiengangsleiter Carlos Albrecht für organisatorische Ideen und Anregungen,

der Gesellschaft der Freunde und Förderer (GFF) der Hochschule für Musik Detmold sowie dem Verband Deutscher Tonmeister (VDT) für die bereitwillige und unkomplizierte großzügige finanzielle Förderung der Diplomarbeit.

Ganz besonderer Dank aber gilt selbstverständlich allen ProbandInnen, die sich Zeit und Ruhe für den Hörvergleich genommen und sich bei den sommerlichen Temperaturen des vergangenen Jahres bereit erklärt haben, uns durch „Hören“ zu unterstützen,

Außerdem sei an dieser Stelle gleichermaßen allen Studierenden und Lehrenden des Erich-Thienhaus-Instituts für das entgegengebrachte Verständnis in Bezug auf räumliche und akustische Einschränkungen während des gesamten Sommersemesters, darüber hinaus ebenfalls den vielen freiwilligen studentischen „Nahrungslieferdiensten“ gedankt.

Zuletzt und in ganz besonderem Maße möchten wir uns jedoch einerseits bei unseren Eltern Annelies und Ulrich Blech und Chin-Hua Huang und Po-Fong Yang für den starken familiären Rückhalt und die jeweils äußerst großzügige (unaufgeforderte) finanzielle Unterstützung, wie auch andererseits ausdrücklich bei Nadine Otterbach und Riky Peßler für die zahlreichen v.a. freundschaftlichen, aber auch fachlichen Gespräche, Anregungen und das akribische und geduldige Korrekturlesen der Arbeit sowie ebenso bei Shu-Han Yen für die bereitwillige Hilfe im Hinblick auf grafische Fragen bedanken. Bezüglich der großen zeitlichen Einschränkungen, die viele Freunde während der Planung und Durchführung der Hörtests erfahren haben, bitten wir um Verständnis.

9 Literaturverzeichnis**Bibliographie und Referenzen:**

- ACKER, Michael; GERICKE, Harald. 1998. *Untersuchung zur Unterscheidbarkeit von analogen und mit 48 kHz und 96 kHz digitalisierten Audiosignalen aus der Sicht des Tonmeisters*. Detmold: Diplomarbeit am Erich-Thienhaus-Institut, Hochschule für Musik Detmold.
- ARATO-BORSI, Eva; POTH, Tamas; FURJES, Andor. 1998. *New Reference Listening Room for Two-Channel and Multichannel Stereophonic*. AES Convention Paper 4732, Presented at the 104th Convention 1998 May 16-19 Amsterdam, The Netherlands.
- AUTORENKOLLEKTIV (Freie Universität Berlin, Universität Bielefeld, Technische Universität Hannover). K.A. *Grundausbildung in Statistik für Wirtschaftswissenschaftler*. Osnabrück: Ruck-Zuck-Druck GmbH.
- BEECK, Torsten. 2002. *Entwicklung eines ABX-Hörtests für eine Multichannel-Audio-Monitoreinheit unter Visual C++*. Hannover: Diplomarbeit an der Fachhochschule Hannover, Fachbereich Elektrotechnik, Studienrichtung Nachrichtentechnik-Kommunikationstechnik.
- BRUHN, Herbert. „Gehör – V. Musikpsychologische Aspekte“, in: *MGG – Die Musik in Geschichte und Gegenwart, Allgemeine Enzyklopädie der Musik*. Sp.1118-1126. 1995: Bärenreiter, Metzler.
- COOK, Perry A. 1999. *Music, Cognition, and Computerized Sound – An Introduction to Psychoacoustics*. Cambridge: The MIT Press.
- dCS-LIMITED. 1999. *User Manual dCS 972*. Saffron Walden, UK.
- EBU European Broadcasting Union. *Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic*. EBU Tech. 3276 – 2nd edition: 1998.
- EBU European Broadcasting Union. *Assessment methods for the subjective evaluation of the quality of sound programme material - Music*. EBU Tech. 3286-E: 1997.
- FRÖHLICH, Werner D. 1994. *dtv-Wörterbuch zur Psychologie*. München: Deutscher Taschenbuch Verlag GmbH&Co.KG.
- GRÜNINGEN, Daniel Ch. von. 1993. *Digitale Signalverarbeitung*. Berlin: vde-Verlag.

- HARRIS, M.N.; KELLY, R.; McLEOD, D.A.; STORY, M.J. 1998. *Effects in High Sample Rate Audio Material*. Presented at: 20th Tonmeistertagung in Karlsruhe, Germany.
- HAWKSFORD, M.O.J. K.A. *Bitstream versus PCM debate for high-density compact disc*. University of Essex, Centre for Audio Research and Engineering.
- ITU Radiocommunication Assembly. 1992. *Test signals and metering to be used on international sound programme connections*. Recommendation ITU-R BS. 645-2: 1986-1990-1992.
- ITU Radiocommunication Assembly. 1994. *Multichannel stereophonic sound system with and without accompanying picture*. Recommendation ITU-R BS. 775-1: 1992-1994.
- ITU Radiocommunication Assembly. 1997. *Methods for the subjective assessment of small impairments in audio systems including Multichannel Sound Systems*. Recommendation ITU-R BS. 1116-1: 1994-1997.
- KAORU, Ashihara; SHOGO, Kiryu. 2001. *Detection threshold for tones above 22 kHz*. AES Convention Paper 5401, Presented at the 110th Convention 2001 May 12-15 Amsterdam, The Netherlands: Audio Engineering Society Reprint.
- LIPSHITZ, Stanley P.; VANDERKOOY, John. 2000. *Why Professional 1-Bit Sigma-Delta Conversion is a Bad Idea*. AES Convention Paper 5188, Presented at the 109th Convention 2000 September 22-25 Los Angeles, California, USA.
- LIPSHITZ, Stanley P.; VANDERKOOY, John. 2001. *Why 1-Bit Sigma-Delta Conversion is Unsuitable for High-Quality Applications*. AES Convention Paper 5395, Presented at the 110th Convention 2001 May 12-15 Amsterdam, The Netherlands.
- LIPSHITZ, Stanley P.; VANDERKOOY, John. 2001. *Towards a Better Understanding of 1-Bit Sigma-Delta Modulators*. AES Convention Paper 5398, Presented at the 110th Convention 2001 May 12-15 Amsterdam, The Netherlands.
- LANGEN, Christian. 1997. *Grundlagen der Analog/Digital-Wandlung*. Vortrag anlässlich des Gemeinschaftsseminars von VDT und AES, 15. und 16.11.1997: 24 bit und 96kHz: Pro und Kontra. Köln.
- MANGER, Daniela L. 1999. *A sound transducer with a flat, flexible diaphragm working with bending waves*. Manger Products, Mellrichstadt, Deutschland.
- MANGER, Josef W. 2000. *Die akustische Realität*. Manger Products, Mellrichstadt, Deutschland.
- MÜLLER, Swen. 2001. „DSD statt PCM: Fort- oder Rückschritt?“, in: *Production Partner 3/2001*: 16-43.
- MÜLLER, Swen. 2002. „Stanley Lipshitz über Digital Audio“, in: *Production Partner 5/2002*: 32-46.

- NISHIGUCHI, Toshiyuki; HAMASAKI, Kimio; IWAKI, Masakazu; ANDO, Akio. NHK Science & Technical Research Laboratories, Tokyo, Japan. 2003. *Perceptual Discrimination between Musical Sounds with and without Very High Frequency Components*. AES Convention Paper 5876, Presented at the 115th Convention 2003 October 10-13 New York, New York.
- PEßLER, Friederike. 2003. *Psychologische und physiologische Aspekte des analytischen Musikhörens*. Detmold: Diplomarbeit am Erich-Thienhaus-Institut, Hochschule für Musik Detmold.
- REEFMAN, Derk; JANSSEN, Erwin. 2002. *Signal processing for Direct Stream Digital – A tutorial for digital Sigma Delta modulation and 1-bit digital audio processing*. (<http://www.superaudiocd.philips.com/Assets/Downloadablefile/wp-2323.pdf>)
- ROEDERER, Juan G. 2000. *Physikalische und psychoakustische Grundlagen der Musik*. Berlin: Springer Verlag.
- SCHMIDT, Robert F. 2001. *Physiologie kompakt*. Berlin: Springer Verlag.
- SCHMIDT, Robert F.; THEWS, Gerhard (Hrsg.). 1990. *Physiologie des Menschen*. Berlin: Springer Verlag.
- SCHÜßLER, Hans Wilhelm. 1994. *Digitale Signalverarbeitung 1*. Berlin: Springer Verlag.
- SILBERNAGEL, Stefan; DESPOPOULOS, Agamemnon. 2001. *Taschenatlas der Physiologie*. Stuttgart: Georg Thieme Verlag.
- SONY Electronics Inc.; PHILIPS Electronics N.V. 1997. *Super Audio Compact Disc - A Technical Proposal*.
- SPITZER, Manfred. 2002. *Musik im Kopf – Hören, Musizieren, Verstehen und Erleben im neuronalen Netzwerk*. Stuttgart: Schattauer GmbH.
- STAHEL, Werner A. 1995. *Statistische Datenanalyse – Eine Einführung für Naturwissenschaftler*. Braunschweig: Vieweg Verlag.
- STEINKE, Gerhard. 1988. *Minimum requirements for reference listening rooms*. AES Convention Paper 2632, Presented at the 84th Convention 1988 March 1-4 Paris, France.
- STORY, Mike. 1997. *A Suggested Explanation For (Some Of) The Audible Differences Between High Sample Rate And Conventional Sample Rate Audio Material*. dCS Ltd Saffron Walden, UK.
- STORY, Mike. 1998. *Timing Errors and Jitter*. dCS Ltd Saffron Walden, UK.
- STORY, Mike; McLEOD, Duncan; REYNOLDS, Martin. 1998. *Resolution, Bits, SNR and Linearity*. dCS Ltd Saffron Walden, UK.
- STOTZ, Dieter. 1994. *Audio- und Videogeräte richtig einmessen und justieren*. München: Franzis'-Verlag.

- STUART, J.Robert. 1996. *DVD: Pre-emphasis for use at 96kHz or 88,2kHz*. J.R.Stuart, for Acoustic Renaissance for Audio, Meridian Audio, Stonehill, Stukeley Meadows, Huntingdon, United Kingdom.
- STUART, J.Robert. 1996. *DVD: A Proposal for the High-Quality Audio Application of High-Density CD Carriers*. J.R.Stuart, for Acoustic Renaissance for Audio, Meridian Audio, Stonehill, Stukeley Meadows, Huntingdon, United Kingdom.
- STUART, J.Robert. K.A. *DVD: Coding High Quality Digital Audio*. J.R.Stuart, for Acoustic Renaissance for Audio, Meridian Audio, Stonehill, Stukeley Meadows, Huntingdon, United Kingdom.
- Ten HOOPEN, G. „Auditive Aufmerksamkeit“, in: Neumann, O.; Sanders, A.F. (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich C, Serie II, Band 2. 115-161*. 1996. Göttingen: Hogrefe.
- TRIMMEL, M. 1996. *Kognitive Psychologie (Skriptum)*. Wien: WUV Universitätsverlag.
- WATKINSON, John. 1995. *The Art of Digital Audio*. Oxford: Focal Press.
- ZIMBARDO, P.G. 1995. *Psychologie*. Berlin: Springer Verlag.
- ZWICKER, Eberhard; FASTL, Hugo. 1999. *Psychoacoustics – Facts and Models*. Heidelberg: Springer Verlag.

Linkographie:

- **ABX-Company:** <http://www.pcavtech.com/abx/>
- **Acoustic Renaissance for Audio:** <http://www.meridian-audio.com/ara>
- **Audio Engineering Society:** <http://www.aes.org>
- **AVM GmbH:** <http://www.avm-audio.com>
- **dCS Limited:** <http://www.dcsLtd.co.uk>
- **European Broadcasting Union:** <http://www.ebu.ch>
- **Emil-Berliner-Studios:** <http://www.emil-berliner-studios.com>
- **Ferromag:** <http://www.ferromag.de>

- **HfM Detmold:** <http://www.hfm-detmold.de>
- **International Telecommunication Union:** <http://www.itu.int/home/>
- **Klotz a.i.s. GmbH:** <http://www.klotz-ais.com>
- **Lake-People electronic GmbH:** <http://www.lake-people.de>
- **Manger Products:** <http://www.manger-msw.com>
- **Media Assistance:** <http://www.media-assistance.com>
- **Merging Technologies:** <http://www.merging.com>
- **Northstar:** <http://www.northstarconsult.nl>
- **Schalltechnik Dr.-Ing. Schoeps GmbH:** <http://www.schoeps.de>
- **Sennheiser electronic GmbH & Co. KG:** <http://www.sennheiser.com>
- **Stax:** <http://www.stax.co.jp>
- **Verband Deutscher Tonmeister:** <http://www.tonmeister.de>

Erklärung

Ich versichere hiermit, dass ich meinen Anteil zu der vorliegenden Gruppenarbeit, die Kapitel 2 und 5, in der gesetzten Frist selbständig verfasst und keine anderen Hilfsmittel als die angegebenen verwendet habe. Alle Stellen der Arbeit, die anderen Werken wörtlich oder sinngemäß entnommen wurden, sind unter Angabe der Quelle als Entlehnung kenntlich gemacht. Das gleiche gilt für die von den auf dem Titelblatt der Arbeit genannten Verfasser gemeinsam verfassten Teile (Kapitel 1, 3 und 6 bis 10).

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Detmold, den 27. Januar 2004

.....
Dominik Blech

Ich versichere hiermit, dass ich meinen Anteil zu der vorliegenden Gruppenarbeit, das Kapitel 4, in der gesetzten Frist selbständig verfasst und keine anderen Hilfsmittel als die angegebenen verwendet habe. Alle Stellen der Arbeit, die anderen Werken wörtlich oder sinngemäß entnommen wurden, sind unter Angabe der Quelle als Entlehnung kenntlich gemacht. Das gleiche gilt für die von den auf dem Titelblatt der Arbeit genannten Verfasser gemeinsam verfassten Teile (Kapitel 1, 3 und 6 bis 10).

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Detmold, den 27. Januar 2004

.....
Min-Chi Yang