

Bachelorarbeit im Studiengang Audiovisuelle Medien  
Hochschule der Medien Stuttgart

# Mehrkanalkodierverfahren für Internetstreaming

-

## Untersuchung von vorhandenen Technologien für das Hörfunkangebot des Bayerischen Rundfunks

vorgelegt am 03.11.2009

von Gerhard Wicho

Matrikelnummer 17790

Erstprüfer: Prof. Oliver Curdt


Zweitprüfer: Dipl. Ing. (FH) Udo Appel

## **Abstract**

Der Bayerische Rundfunk bietet seit einigen Jahren regelmäßig Hörfunksendungen im Mehrkanalformat an. Verbreitungswege dafür sind bisher die digitale Satellitenübertragung und zum Teil auch das digitale Kabelnetz, weitere Distributionsarten wurden bislang nur in Probeläufen getestet. In dieser Arbeit werden vorhandene Technologien untersucht, die für das Internetstreaming in Frage kommen, mit dem Ziel, anhand von Hörtests die Akzeptanz der Audioqualität dieser Mehrkanalkodierverfahren für potenzielle Programminhalte des Bayerischen Rundfunks festzustellen.

# Inhaltsverzeichnis

<b>Abstract</b> .....	<b>2</b>
<b>Inhaltsverzeichnis</b> .....	<b>3</b>
<b>Abbildungsverzeichnis</b> .....	<b>5</b>
<b>1. Einleitung</b> .....	<b>6</b>
<b>2. Begriffserläuterungen</b> .....	<b>7</b>
2.1 Mehrkanalton .....	7
2.2 Streaming .....	8
2.3 Kodierverfahren und Bitratenreduktion.....	9
<b>3. Bandbreitenbedarf und Audiokodierverfahren</b> .....	<b>10</b>
3.1 Datenraten bei der digitalen Informationsübertragung im Hörfunkbereich des Bayerischen Rundfunks.....	10
3.1.1 DVB-S .....	10
3.1.2 DVB-C .....	11
3.1.3 Digitalradio .....	11
3.1.4 Internetlivestream .....	12
3.2 Ansätze von Audiokodierverfahren zur Bitratenreduktion .....	13
3.2.1 Verlustlose Kodierung .....	13
3.2.2 Verlustbehaftete Kodierung.....	13
3.2.3 Räumliche Parametrische Kodierung .....	15
3.2.4 Wahrnehmbare Qualitätsverluste bei verlustbehafteter Audiokodierung.....	17
<b>4. Grundlagen der Streamingtechnologie</b> .....	<b>18</b>
4.1 Das Client-Server Modell .....	18
4.2 Anforderungen an Streamingsysteme.....	18
4.3 Streamingfähige Protokolle.....	19
<b>5. Erläuterung der MUSHRA Methode</b> .....	<b>21</b>
5.1 Einsatzgebiet.....	21
5.2 Voraussetzungen für die Durchführung .....	21
5.2.1 Reihenfolge der Materialien .....	21
5.2.2 Einführende Trainingssession .....	22
5.2.3 Abhörbedingungen.....	22

5.2.4 Auswahl des Audiomaterials.....	22
5.2.5 Auswahl der Testpersonen .....	23
5.2.6 Ausschlusskriterien in den gesammelten Bewertungen.....	23
5.2.7 Statistische Analyse.....	23
5.2.8 Der Wilcoxon-Vorzeichen-Rang-Test als ergänzendes Mittel zur statistischen Analyse .....	25
<b>6. Durchführung der Hörtests .....</b>	<b>26</b>
6.1 Auswahl der Audiobeispiele.....	26
6.2 Formate im Vergleich.....	27
6.2.1 MP3 Surround .....	27
6.2.2 Windows Media Audio .....	27
6.2.3 Ogg Vorbis .....	28
6.2.4 Dolby Digital .....	28
	
6.3 Eingesetzte Software bei der Vorbereitung der Hörtests.....	29
6.4 Auswahl der Testpersonen .....	31
6.5 Testsoftware .....	31
6.6 Testumgebung.....	33
6.6.1 Raum .....	33
6.6.2 Eingesetzte Technik.....	34
6.7 Trainingsession .....	35
6.8 Durchführungszeitraum und Dauer der Tests .....	35
<b>7. Auswertung und grafische Aufbereitung .....</b>	<b>36</b>
7.1 Boxplot.....	36
7.2 Darstellung der Mittelwerte und Konfidenzintervalle nach Kodierverfahren und Hörbeispielen aufgeteilt .....	40
7.2.1 Alle Kodierverfahren über alle Hörbeispiele .....	41
7.2.2 Mittelwerte und Konfidenzintervalle aller Hörbeispiele separat für jedes Kodierverfahren .....	43
7.2.3 Ergänzende Darstellung der Mittelwerte und Konfidenzintervalle aller Kodierverfahren separat für jedes Hörbeispiel .....	50
<b>8. Fazit .....</b>	<b>58</b>
<b>Literaturverzeichnis .....</b>	<b>59</b>
<b>Erklärung .....</b>	<b>60</b>

## **Abbildungsverzeichnis**

Abb. 1 Aufstellung 3/2 Mehrkanalsystem .....	8
aus ITU-R B.S. 775-2 (siehe Literaturverzeichnis)	
Abb. 2 Prizipieller Aufbau des Client Server Modells bei Streaming .....	18
entspricht in etwa der Grafik in Lee (2005) S. 12 (siehe Literaturverzeichnis)	
Abb. 3 Bedienoberfläche von STEP während eines Tests .....	32
Screenshot der Testsoftware während einer Evaluierung	
Abb. 4 Skizze Einrichtung DSR5 .....	33
Grundriss und Einrichtungsplan des DSR5 (Digitaler Schneiderraum 5) im Funkhaus München	
Abb. 5 Blick auf Arbeitsplatz und Frontlautsprecher DSR5 .....	34
Foto mit Blick über den Sweetspot und den Arbeitsplatz auf die Frontlautsprecher im DSR5	

# 1. Einleitung

Seit 2005 finden im Programm BR Klassik (bis 30. September 2009 Bayern 4 Klassik) regelmäßig mehrkanalige Liveausstrahlungen aus der Philharmonie am Gasteig in München, aus dem Herkulesaal der Residenz in München, sowie von den Bayreuther Festspielen statt. In Dolby Digital kodiert (mit einer Datenrate von 448 kbps) erreicht das Programm über DVB-S und teilweise über DVB-C die Hörer mit einer Tonqualität, wie man sie auch häufig bei DVD-Video antrifft.

Seit der Einführung der DVD Mitte der Neunziger Jahre gibt es einen Trend hin zu mehrkanaligen Tonmischungen und in vielen Haushalten findet sich mittlerweile ein System zur Wiedergabe mehrkanaligen Audiomaterials. Von immer größer werdender Bedeutung ist diese Entwicklung, seit die Fernsehübertragung zunehmend auf digitaler Ebene stattfindet und verstärkt Spielfilme oder Sportsendungen im Mehrkanalformat angeboten werden.

Zudem verfügen viele Haushalte heutzutage über einen breitbandigen Internetzugang und multimediafähige Rechner, die entsprechende Audioanschlüsse standardmäßig integriert haben. Um Hörern der Klassikwelle des Bayerischen Rundfunks, die nicht über die nötige Hardware verfügen um digitale Satelliten- oder Kabelempfang nutzen zu können, den Hörgenuss einer mehrkanaligen Konzertübertragung zu ermöglichen, könnte daher ein entsprechendes Internetstreaming eine attraktive Erweiterung des Programmangebots des Bayerischen Rundfunks darstellen. Nicht nur im Klassiksektor, sondern auch bei der Ausstrahlung von Jazz- oder Popkonzerten sowie Hörspiel- oder Featureproduktionen würde dabei das weltweit empfangbare Hörfunkangebot weiter bereichert werden.

In dieser Arbeit werde ich es mir zur Aufgabe machen, die Qualität von verschiedenen Verfahren zur Mehrkanalkodierung zu vergleichen, um feststellen zu können, welche sich davon besonders für die Distribution über das World Wide Web anbieten. Anhand von Hörtests, nach einer Empfehlung der ITU durchgeführt, möchte ich signifikante Unterschiede untereinander sowie zum Original ausmachen.

## 2. Begriffserläuterungen

### 2.1 Mehrkanalton

Wenn man von Mehrkanalton spricht, so ist in der Regel eine Erweiterung des Stereotons gemeint, die unter Verwendung zusätzlicher Kanäle und Lautsprecher den Hörer mit einem allseitigen Klangfeld umgibt (häufig auch als Mehrkanalstereofonie bezeichnet). Es sind mehrere Aufstellungsarten der Lautsprecher möglich, die gängigste jedoch ist die Verwendung einer 3/2 Anordnung mit drei Lautsprechern vor (L für Left, C für Center, R für Right, also links, mittig und rechts) sowie zwei Lautsprechern seitlich hinter dem optimalen Abhörplatz, dem sogenannten Sweetspot (LS für Left Surround - links hinten und RS für Right Surround - rechts hinten). Optional kann ein weiterer Kanal (LFE für Low Frequency Effect) für die Wiedergabe tieffrequenter Signale die Anordnung ergänzen.

Den De-facto-Standard für Mehrkanaltonsysteme stellt eine Empfehlung der ITU (International Telecommunication Union) dar, die in ihrer Publikation ITU-R BS.775 die Aufstellung von 3/2 sowie abweichenden Systemen definiert, z. B. 2/2 oder 3/4, auf die an dieser Stelle nicht weiter eingegangen wird.

Für die 3/2 Anordnung wird dabei ein Kreis beschrieben, auf dem alle Lautsprecher aufgestellt werden und in dessen Mittelpunkt der Sweetspot liegt. Der linke und rechte Lautsprecher bilden mit dem Sweetspot ein gleichseitiges Dreieck. Zwischen diesen platziert man den Centerlautsprecher. Die Verbindung zwischen C und dem Sweetspot kann als Mittelachse angesehen werden, von der aus sich in einem Winkel von 100 bis 120 Grad in beiden Richtungen die Surroundlautsprecher LS und RS befinden. Es wird zudem empfohlen, dass die Höhe von L, C und R etwa bei 1,20 m liegt, bei LS und RS sollte sie mindestens 1,20 m bei einem Neigungswinkel bis zu 15 Grad nach unten betragen, wobei immer eine Ausrichtung zum Sweetspot hin erforderlich ist.<sup>1</sup>

Optimalerweise sollten diese Kanäle über Lautsprecher wiedergegeben werden, die den Frequenzbereich von 20 - 20000Hz möglichst linear reproduzieren. Diese Bedingung ist aber meist nur im Studio erfüllt, im Heimbereich findet man oft Systeme die den Bassbereich nur eingeschränkt übertragen. Zusätzlich zu den fünf Kanälen wird häufig ein LFE Kanal eingesetzt, der nur den Frequenzbereich von 20 Hz - ca. 80 Hz bis 120 Hz wiedergibt. Daraus ergibt sich dann die Bezeichnung 5.1 für die Wiedergabe über fünf Full Range Lautsprecher und einen bandbreitenbegrenzten Subwoofer. Neben dem LFE Kanal wird über einen Sub-

---

<sup>1</sup> Vgl. ITU-R B.S. 775-2 S. 2 ff

woofer häufig auch der tiefe Frequenzbereich der 3/2 Anordnung übertragen, da viele Systeme dem Anspruch nicht gerecht werden, den kompletten Bereich von 20 - 20000 Hz übertragen zu können und daher keine Full Range Systeme sind. Die größten Probleme ergeben sich dabei meist im Bassbereich, so dass dieser im Rahmen eines Bass-Managements über eine aktive Mehrkanal-Frequenzweiche abgetrennt wird und dem Subwoofer gemeinsam mit dem LFE Kanal zugeführt wird. Die Aufstellung des Subwoofers ist relativ frei, da die Signale im tieffrequenten Bereich nur schwer zu orten sind.<sup>2</sup>

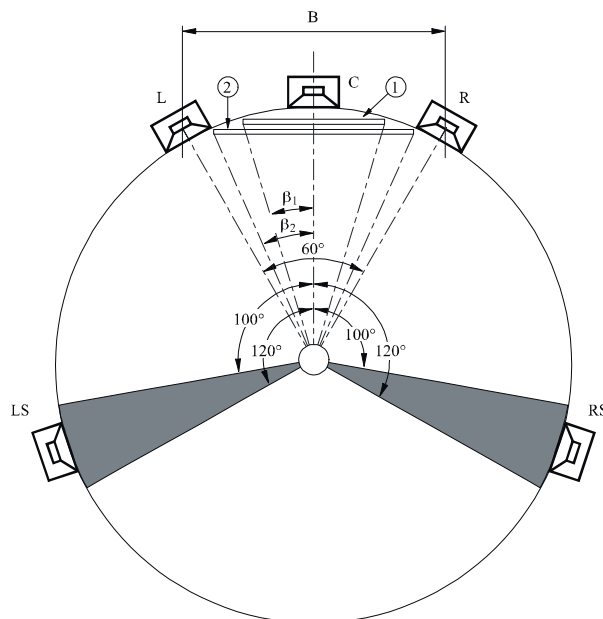


Abb. 1 Aufstellung 3/2 Mehrkanalsystem

B: beschreibt die Basisbreite, also den Abstand zwischen L und R, dadurch auch den Abstand der Lautsprecher zum Sweetspot

① und ②: bezeichnen mögliche Positionen von Bildschirmen, wenn Mehrkanalton in Verbindung mit Bewegtbild genutzt wird

Die Winkel  $2\beta_1$  und  $2\beta_2$  betragen  $33^\circ$  bzw.  $48^\circ$ , daraus ergeben sich die Abstände von  $3 \times H$  zu Bildschirm 1 (HDTV Referenz Abstand) und  $2 \times H$  zu Bildschirm 2, wobei H die Höhe des Bildschirms beschreibt.

## 2.2 Streaming

Als Streaming bezeichnet man die Übertragung von Multimediatechniken wie Audio- oder Videoinhalten, wobei ein Abspeichern der Daten nicht erforderlich ist, da sie bereits während der Übertragung wiedergegeben werden können. Somit grenzt sich Streaming von Downloadangeboten ab, bei welchen ein Abspielen erst nach dem vollständigen lokalen Speichern einer Datei möglich ist.

Man unterscheidet zwei Arten von Streaming. Zum einen das On-Demand-Streaming, bei dem Mediafiles auf einem Streamingserver hinterlegt werden und die auf Anfrage von Usern abrufbar sind. Die Daten können sofort in der entsprechenden Abspielsoftware gesehen bzw. gehört werden; während der Wiedergabe kann vor- und zurückgespult sowie unterbrochen werden. Zum anderen das Livestreaming, bei dem Medieninhalte anhand eines Hard- oder

<sup>2</sup> Vgl. Slavik und Weinzierl (2008) S. 645



Softwareencoders in digitaler Form mit geringer zeitlicher Verzögerung zur Verfügung gestellt werden und von einem Player mit entsprechendem Decoder wiedergegeben werden können. So lässt sich z. B. ein Radioprogramm parallel zur Ausstrahlung über UKW auch über das Internet übertragen. Ein Vor- oder Zurückspulen ist nicht möglich, es wird immer nur die aktuelle Information, beispielsweise ein Radioprogramm, weitergegeben.<sup>3</sup>

### **2.3 Kodierverfahren und Bitratenreduktion**

Verschiedentliche Kodierverfahren können zum Einsatz kommen um die Datenmenge, etwa von Audio-, Video- oder Bilddaten, für die Speicherung oder Übertragung zu verringern. Der sogenannte Encoder setzt am Anfang des Prozesses das Ausgangssignal nach einer bestimmten Rechenvorschrift um und stellt die digitale Information in einem bestimmten Code dar. Der Decoder entschlüsselt diesen am Ende der Kette und versucht die vom Encoder erstellten Daten wieder möglichst in ihren Originalzustand für die Wiedergabe zurückzuwandeln.

Dabei gibt es einmal den Ansatz der verlustfreien Kodierung, deren Ziel es ist, redundante, also verzichtbare Information aus den Dateien zu entfernen, die Medieninhalte hinterher aber wieder unverfälscht rekonstruieren zu können. Die Daten können dabei mehrfach encodiert und ohne Qualitätsverlust wieder decodiert werden.

Bei verlustbehafteten Kodierverfahren wird das Signal analysiert und neu quantisiert. Vermeintlich verzichtbare Informationen werden dabei gezielt ausgelassen und mögliche Qualitätsverluste in Kauf genommen. Je nach Anforderung, ausgewähltem Algorithmus und Encoder ist diese Veränderung des Materials mehr oder weniger auffällig. Im Allgemeinen lässt sich mit der verlustbehafteten Kodierung eine deutlich stärkere Bitratenreduktion erreichen (durchschnittlich Faktor 10 - 50) als mit verlustfreien Verfahren (bis zu Faktor, 3 abhängig vom Ausgangsmaterial). Eine Kaskadierung, die mehrfach nacheinander stattfindende Anwendung verlustbehafteter Kodierverfahren auf das gleiche Material, kann zu inakzeptabel fehlerhaften Ergebnissen führen.<sup>4</sup> Dennoch ist Bitratenreduktion auf Kosten der Qualität in vielen Bereichen unumgänglich. So sind beispielsweise bei der Verbreitung von Audio und Video über das Internet die Grenzen der Datenrate durch die verfügbare Übertragungsbandbreite vorgegeben. Bei der Erstellung von DVDs sind die Auslesegeschwindigkeit, welche im DVD Standard festgelegt ist, als auch die Kapazität für die Speicherung von Audio- oder Bilddaten nicht unbeschränkt, so dass häufig im letzten Schritt des Bearbeitungsprozesses eine Datenreduktion stattfindet.

---

3 Vgl. Bayerischer Rundfunk (2000) S. 265

4 Vgl. Lerch (2008) S. 849 f

### 3. Bandbreitenbedarf und Audiokodierverfahren

Je nach Anwendungsfall muss auf digitaler Ebene eine Datenreduktion durchgeführt werden. Bei der Übertragung ist ebenso wie bei der Speicherung von Audio- oder Videodaten, meist eine Reduzierung der Datenrate notwendig um innerhalb der verfügbaren Bandbreite oder in den Grenzen des verfügbaren Speicherplatzes eine möglichst hohe Qualität zu gewährleisten.

#### 3.1 Datenraten bei der digitalen Informationsübertragung im Hörfunkbereich des Bayerischen Rundfunks

Das Programm des Bayerischen Rundfunks ist auf unterschiedlichsten Wegen empfangbar. Die wohl noch am weitesten verbreiteten Empfangsgeräte sind analoger Art und werden durch die UKW-Sender versorgt. Daneben ist aber auch ein Empfang auf digitaler Basis möglich, der immer mehr an Bedeutung gewinnt. Die digitale Verbreitung wird auf vier verschiedene Arten abgewickelt, nämlich über Satellit (DVB-S), als Einspeisung ins Kabelnetz (DVB-C), im Digitalradio (ehemals DAB) und als Stream über das Internet. Während bei der Übertragung über Satellit, Kabel und Digitalradio mit einer festen zur Verfügung stehenden Bandbreite gerechnet werden kann, schwanken die Zahlen beim Internetstreaming sehr stark. Abhängig von den technischen Einrichtungen des Internet Service Providers, der Art der vorhandenen Übertragungsleitungen sowie von der individuellen Wahl des Internetanschlusses kann die Übertragungsgeschwindigkeit zwischen 54 kbps (analoges Modem) und mehreren tausend kbps (Breitbandanschluss) liegen.

##### 3.1.1 DVB-S

Für die Ausstrahlung der Programme über DVB-S wird generell auf das Musicamformat (MPEG-1 Audio Layer 2) zurückgegriffen. Zusätzlich dazu gibt es für BR Klassik ein Angebot in Dolby Digital. Die Datenraten für den Empfang der Programme (Satellit Astra 1H) betragen derzeit:

Programm	Datenrate in kbps	Format
Bayern 1	320	Musicam
Bayern 2	320	Musicam
Bayern 3	320	Musicam

BR Klassik	320	Musicam
BR Klassik	448	Dolby Digital 2.0
Ausgewählte Mehrkanaltonsendungen	448	Dolby Digital 5.1
B 5 aktuell (mono)	128	Musicam
B 5 plus (mono)	128	Musicam
Bayern plus	320	Musicam
on3	320	Musicam
BR Verkehr (mono)	128	Musicam

Die Ausstrahlung erfolgt generell in stereo, die Nachrichten- und Verkehrsangebote (B5 aktuell, B5 plus und BR Verkehr) werden mono angeboten. Dadurch kann auch die Bitrate reduziert werden, ohne deutliche Qualitätseinbußen hinnehmen zu müssen.<sup>5</sup>

### 3.1.2 DVB-C

Die Einspeisung ins Kabelnetz ist regional unterschiedlich und hängt vom jeweiligen Kabelnetzbetreiber ab. Über Programmvierfalt und Qualität lassen sich daher keine allgemeinen Aussagen treffen. Da aber häufig das Angebot, welches auch über Satellit empfangbar ist, ins Kabelnetz übernommen wird, stehen vielerorts ebenso viele Programme zur Verfügung.

### 3.1.3 Digitalradio

Für die Nutzer von Digitalradio werden folgende BR-Hörfunkprogramme im Musicamformat (MPEG-1 Audio Layer 2) ausgestrahlt:

Programm	Datenrate in kbps	Format
Bayern 2 plus (größtenteils identisch mit dem Programm von Bayern 2)	128	Musicam
BR Klassik	192	Musicam
B 5 plus (mono)	96	Musicam
Bayern plus	128	Musicam
on3	128	Musicam
BR Verkehr (mono)	64	Musicam

Im Digitalradio variiert die Datenrate mehr als bei DVB-S. Da aus frequenztechnischen Grün-

<sup>5</sup> Online: <http://www.br-online.de/unternehmen/technik/rundfunktechnik-DID1200666073148/rundfunktechnik-radio-satellit-ID671202493936395040.xml> (Stand: 01.09.2009)

den im Digitalradio die Programme nicht separat übertragen werden, sondern in Programmpaketen (auch Ensembles genannt) mit einer festgelegten Bandbreite zusammengefasst werden, müssen die Datenraten für die einzelnen Programme entsprechend verteilt werden. Die aufgeführten Programme werden in dem landesweiten Paket gemeinsam mit den Sendern Rock Antenne und Radio Galaxy in ganz Bayern verbreitet.

Die höchste Datenrate wurde dem Klassikprogramm zugewiesen, da die Audioqualität bei klassischer Musik am ehesten unter geringen Bitraten leidet. B 5 plus und BR Verkehr werden wiederum nur mono ausgestrahlt um Bandbreite zu sparen, BR Verkehr dabei sogar nur mit einer Rate von 64 kbps, da es nicht als durchhörbares Programm konzipiert ist und die Qualität für die Vermittlung der Information ausreicht.

In München wird zudem ein regional empfangbares Testensemble des Bayerischen Rundfunks ausgesendet, in dem Bayern 1, Bayern 2 plus und Bayern 3 bei einer Datenrate von 192 kbps enthalten sind.<sup>6</sup>

Zu Testzwecken wurde am 28. Dezember 2008 einen Tag lang vorproduziertes Mehrkanaltonmaterial über das Digitalradio verbreitet. Ein Datenformat, welches für Mehrkanalübertragung via Digitalradio in Zukunft in Frage kommen könnte und an diesem Tag auch ausgestrahlt wurde, ist MPEG Surround. In dem Datenstrom wurde ein Stereosignal weitergegeben sowie Zusatzinformationen, die aus der ursprünglichen Mehrkanalmischung errechnet wurden und anhand derer der Dekoder im Idealfall die anfängliche Mehrkanalfassung generieren kann.

### 3.1.4 Internetlivestream

Für den Radioempfang über das Internet stellt der Bayerische Rundfunk derzeit jeweils vier Varianten aller empfangbaren Programme (Bayern 1, Bayern 2, Bayern 3, BR Klassik, B5 aktuell, B5 plus, Bayern plus, on3) auf seiner Homepage [www.br-online.de](http://www.br-online.de) zur Verfügung:

Programm	Datenrate in kbps	Format
alle gestreamten Programme sind in vier Varianten verfügbar	56	MP3
	128	
	48	Windows Media Audio
	128	

<sup>6</sup> Online: <http://www.ukwv.de/sender-tabelle/index.html> (Stand: 01.09.2009)

Für die acht Programme werden zwei verschiedene Formate (MP3 und Windows Media) mit je zwei verschiedenen Bitraten angeboten um sowohl Menschen mit einem langsameren Internetzugang über ein analoges Modem, als auch Hörer mit einem Breitbandanschluss versorgen zu können. Mit 128 kbps geht man dabei allerdings noch nicht an die Grenzen und höhere Datenraten wären durchaus möglich bei immer schnelleren DSL Angeboten. Das ebenfalls weit verbreitete RealAudio Format der Firma RealNetworks soll ARD-weit für Internetstreamingangebote nicht mehr unterstützt werden und findet daher an dieser Stelle keine Beachtung.<sup>7</sup>

## **3.2 Ansätze von Audiokodierverfahren zur Bitratenreduktion**

### **3.2.1 Verlustlose Kodierung**

Wie bereits in Punkt 2.3 erwähnt, gibt es Verfahren, die es ermöglichen die Datenrate ohne Qualitätsminderung zu verringern. Im Bereich der Audiotechnik sind dies unter anderem FLAC (Free Lossless Audio Codec), MPEG-4 ALS (Audio Lossless), Windows Media Audio Lossless oder MLP (Meridian Lossless Packing). Letzteres wird bei der Erstellung von DVD Audio angewendet, wenn die Auslesegeschwindigkeit der DVD von maximal 10,08 Mbps bei der Wiedergabe des Tonmaterials überschritten werden würde. Dies ist bei einem sechskanaligen Signal der Fall, das mit 96 kHz bei einer Bittiefe von 24 Bit abgetastet wurde. Die erforderliche Datenrate von 13,5 Mbps würde vom Player nicht unterstützt werden. Noch höhere Abtastraten, z. B. von 192 kHz, würden entsprechende Konsequenzen nach sich ziehen. Bei der Vorbereitung der Hörtests wurde ich mit dieser Art von Kodierung konfrontiert, da die verlustlose Kodierung ansonsten für diese Arbeit aber keine weitere Rolle spielt, wird an dieser Stelle nicht weiter darauf eingegangen.

### **3.2.2 Verlustbehaftete Kodierung**

Um ein Hörfunkprogramm digital übertragen zu können, ist es unumgänglich, die Datenrate so drastisch zu reduzieren, dass Qualitätsverluste nicht vermeidbar sind. Ein Stereosignal einer CD würde für die bitgenaue Übertragung 1411,2 kbps (2 x 44,1 kHz x 16 Bit) erfordern, ein sechskanaliges Signal bei 96 kHz und 24 Bit, wie es häufig auf einer DVD-Audio zu finden ist, sogar 13824 kbps. Sollte ein solches Signal beispielsweise auf BR Klassik über das Digitalradio übertragen werden, wo eine Bandbreite von 192 kbps dafür bereitgehalten

<sup>7</sup> Online: <http://www.br-online.de/unternehmen/technik/rundfunktechnik-DID1200666073148/rundfunktechnik-radio-livestreams-ID671202493946869045.xml> (Stand: 01.09.2009)

wird, entspricht das einer Verkleinerung im Verhältnis 1:72. Da diese Einsparung an Information nicht ohne Qualitätsminderung vor sich geht, sind verlustbehaftete Kodierverfahren für die Distribution beim Endverbraucher und die Speicherung geeignet, nicht aber für professionelle Anwendungen, denn bei der Weiterverarbeitung des Signals und einer eventuellen mehrfachen Kodierung, werden die Qualitätsverluste schnell deutlich hörbar.

Nun gilt es trotz allem das Material bei der Datenreduktion so wenig wie möglich zu beeinträchtigen. Die meisten gängigen Verfahren basieren dabei auf der Analyse und Erkennung nicht relevanter Signalanteile aufgrund von psychoakustischen Modellen. Diese wahrnehmungsangepasste Kodierung macht sich vor allem Verdeckungseffekte zu Nutze. In Untersuchungen wurde festgestellt, dass laute Signalanteile gleichzeitig auftretende leisere verdecken können (Simultanverdeckung), man spricht dann von Maskierung. Diese Maskierung findet in gewissen Grenzen auch dann statt, wenn leisere Signalanteile vor oder nach einem lauterem Signal auftreten (Vor- bzw. Nachverdeckung).

Zunächst muss eine Analyse durchgeführt werden um diese irrelevante Information auszumachen. Dazu wird das Audiomaterial zeitlich segmentiert, wobei sich die entstandenen sogenannten Fenster überlappen sowie in Frequenzbänder mittels einer Filterbank oder Transformation unterteilt, deren Design maßgeblich die Qualität des Verfahrens bestimmt. An dieser Stelle ist es von erheblicher Bedeutung, dass das Signal möglichst originalgetreu rekonstruiert werden kann und die entstandenen Subbänder gut voneinander getrennt sind. Dies ist bei dem häufig verwendeten MDCT (Modified Discrete Cosine Transform) Verfahren der Fall, das außerdem eine effiziente Lösung für die zu berechnenden Subbandkoeffizienten darstellt. Abhängig vom Ausgangsmaterial sollte die Auflösung im Frequenz- sowie im Zeitbereich variieren. Für sich zeitlich stark verändernde Audiosignale werden kurze Blöcke mit geringerer Frequenzauflösung angestrebt, bei länger klingenden Signalen können diese bei gleichzeitiger Erhöhung der Frequenzauflösung entsprechend größer sein.

Das zugrunde liegende psychoakustische Modell bestimmt dann im jeweiligen Frequenzband, welche Signalanteile maskiert werden und somit irrelevant also vernachlässigbar sind (Signal-To-Mask-Ratio bzw. SMR). Das Frequenzspektrum wird in Bänder unterteilt, die sogenannten Frequenzgruppen, deren Breite mit steigender Frequenz nichtlinear zunimmt. Eine Einteilung des Spektrums und somit der Tonheit erfolgt dabei in der Pseudoeinheit Bark, wobei sich die Anzahl der Bark mit steigender Frequenz erhöht. Entscheidend bei der Festlegung der Maskierungsschwelle ist auch, ob es sich um einen eher sinusförmigen oder einen rauschhaften Maskierer handelt. Die Vorschrift, in welcher Weise ein Signal andere Signalanteile

verdeckt, wird als Prototypfunktion bezeichnet und die sich ergebende Verdeckungsschwelle durch Faltung mit dem Barkspektrum ergibt das resultierende akustische Ereignis bei einer Simultanverdeckung. Der Verlauf der Ruhehörschwelle, dadurch auch die Handhabung von Vor- und Nachverdeckungseffekten, kann in den jeweils verwendeten Modellen voneinander abweichen und daher zu unterschiedlichen Ergebnissen führen.

Der nächste Schritt wäre nun die Bitallokation, die die Neuquantisierung aufbauend auf den vom psychoakustischen Modell ermittelten Daten steuert. Zwei Kriterien sind bei diesem Vorgang in Betracht zu ziehen, zum einen das bei diesem Schritt unvermeidbare Quantisierungsrauschen in den einzelnen Frequenzbändern möglichst unterhalb der Maskierungsschwelle zu halten und zum anderen die vorgegebene Bitrate zu erreichen. Eine anschließende Entropiekodierung, bei der häufiger auftretende Werte mit einer geringeren Bitzahl kodiert werden als selten auftretende, hilft dabei die angestrebte Datenrate auch tatsächlich einzuhalten. Eine Abweichung ist allerdings nicht immer zu vermeiden, denn die Entropiekodierung ist stark abhängig von der Beschaffenheit des Signals.

Weitere Datenersparnis kann durch das sogenannte Joint Channel Coding erreicht werden. Einerseits können Abhängigkeiten zwischen den Eingangskanälen in eine gemeinsame Form gebracht werden und dadurch deutlich effizienter dargestellt werden als die Einzelkanäle (z. B. Summen- und Differenzbildung bei der Mitte/Seite-Matrizierung zweier Kanäle). Andererseits wurde nachgewiesen, dass bei hohen Frequenzen die Lokalisation nicht mehr eindeutig ist und hauptsächlich über Pegeldifferenzen wahrgenommen wird. Die bei niedrigen Bitraten eingesetzte verlustbehaftete Intensitätskodierung verzichtet daher darauf, höhere Frequenzbänder einzeln zu berechnen, kombiniert diese und fügt ihnen Seiteninformationen hinzu, welche die Pegelunterschiede repräsentieren.<sup>8</sup>

### 3.2.3 Räumliche Parametrische Kodierung

Für die Kodierung von mehrkanaligem Audiomaterial gibt es nun weitere, allerdings verlustbehaftete Ansätze, die dem Joint Channel Coding ähnlich sind, die Möglichkeiten jedoch erweitern. Im Besonderen möchte ich das Binaural Cue Coding (BCC) und das Spatial Audio Coding (SAC) kurz erläutern. Diese Verfahren basieren auf Erkenntnissen über die räumliche Wahrnehmung aus der Psychoakustik. Durch Analyse des Materials in den bereits erwähnten Frequenzgruppen können verschiedene Parameter berechnet werden, die man dann als Seiteninformation einem Summensignal der einzelnen Kanäle beim Encodieren hinzufügen

---

8 Vgl. Lerch (2008) S. 857 ff

kann. Der Decoder gibt dann entweder das Summensignal wieder (Abwärtskompatibilität) oder, wenn dies gewünscht ist, synthetisiert aus den Seiteninformationen die ursprünglichen Kanäle anhand der Parameter.

Beim BCC baut man darauf auf, dass die binaurale Wahrnehmung im Wesentlichen von drei Faktoren abhängt, den interauralen Pegelunterschieden (ILD für Interaural Level Difference), den interauralen Zeitdifferenzen (ITD für Interaural Time Difference) und der interauralen Korrelation (IC für Interaural Correlation), zusammen auch interaurale Cues genannt. Übertragen auf ein Audiosystem spricht man dann bei der Kodierung von Inter-Channel Level Difference (ICLD), Inter-Channel Time Difference (ICTD) und Inter-Channel Correlation (ICC). Das Signal wird wieder in die einzelnen Frequenzgruppen aufgeteilt und für jedes Teilband ein ICLD und ein ICTD bezüglich eines Referenzkanals (die Berechnung für jedes einzelne Kanalpaar wäre eigentlich möglich, aber wird als zu aufwändig betrachtet) übertragen, um die Richtungsinformation im Decoder wieder rekonstruieren zu können. Für die Lateralisation sind dabei ICLDs für Frequenzen oberhalb von ca. 1 -1,5 kHz von größerer Bedeutung als ICTDs, für den unteren Frequenzbereich gilt entsprechend, dass ICTDs mehr Einfluss haben. Zudem wird für jedes Teilband die Kohärenz zwischen den beiden Kanälen mit der höchsten Energie berechnet und weiter gegeben. Das Summensignal, auf das die Parameter angewendet werden, ist ein aus allen Kanälen gebildetes Monosignal, wobei berücksichtigt wird, dass die Energie der summierten Signalanteile in etwa an die Energie in den Eingangskanälen angeglichen wird, so dass es nicht zur Verstärkung oder Abschwächung einzelner Anteile bei der Summenbildung kommt. Aus der Summe und den Cues wird dann im Decoder die Rauminformation wieder synthetisiert um daraus das Originalsignal möglichst unverfälscht wiederherzustellen.<sup>9</sup>

Ähnlich verläuft das SAC; bei der Analyse des Signals werden dabei die Parameter Channel-Level Difference (CLD - analog zu den ICLD), Inter-Channel Correlation (ICC), Channel-Prediction Coefficient (CPC - zur Prädiktion der Ausgangssignale aus den Eingangssignalen) und Prediction Errors (auch Residual Signal - sie beschreiben den Unterschied zwischen der parametrischen Beschreibung der Wellenform und der tatsächlichen Wellenform) erzeugt und als Zusatzinformation zur gebildeten Summe übertragen. Im Gegensatz zum BCC kann diese jedoch mehrkanalig, z. B. stereo sein.<sup>10</sup>

---

9 Vgl. Faller und Baumgarte (2003) S. 520 ff

10 Vgl. Herre (2005) S. 2 ff



### 3.2.4 Wahrnehmbare Qualitätsverluste bei verlustbehafteter Audiokodierung

Die Folgen von verlustbehafteter Audiokodierung können je nach Verfahren und verwendeter Bitrate hörbar sein, müssen aber nicht immer zwingend in gleicher Weise auftreten. Typische Artefakte, die eine Verschlechterung des Audiomaterials nach sich ziehen, sind Vorecho, Verschmierungen, Aliasing (Zwitschern) und fehlerhafte Höhenwiedergabe, Granularrauschen oder falsche räumliche Abbildung.

Ein Vorecho tritt auf, wenn ein transientes Signal, z. B. bei der Aufnahme eines Schlaginstruments, mit einer zu geringen Auflösung im Zeitbereich auf eine stille Passage folgt. Das zwangsläufig auftretende Rauschen bei der Neuquantisierung des Signals verteilt sich auf die komplette Blocklänge und kann bei entsprechend langen Blöcken in einem leisen Abschnitt vor dem eigentlichen Hörereignis als Vorecho hörbar werden. Bei entsprechend kurzen Blocklängen verhindert dies der Vorverdeckungseffekt.

Verschmierungen werden erzeugt, wenn hohe Frequenzanteile eines transienten Signals wegen zu grober Quantisierung zu ungenau analysiert und daher nicht mehr klar definiert dargestellt werden.

Aliaseffekte können dadurch entstehen, dass die Grenzfrequenz des Tiefpassfilters vor der eigentlichen Kodierung zu hoch angesetzt wird, was Artefakte in den höheren Frequenzbereichen nach sich zieht und als Zwitschern hörbar wird. Ist diese andererseits zu tief eingestellt, kann es passieren, dass in den Höhen Signalanteile fehlen, was ebenfalls als störend empfunden werden kann. Dies ist allerdings stark von der Art des Signals und dem Verlauf des Spektrums abhängig.

Des Weiteren kann das durch den Quantisierungsvorgang hinzugefügte Rauschen von Block zu Block stark variieren. Wenn diese Modulation des Granularrauschens im Ausgangssignal zu auffällig ist, erzeugt sie den subjektiven Eindruck der Rauigkeit.

Bei der Kodierung von Rauminformationen können außerdem zeitliche Verschiebungen aufgrund von Schwächen des Systems auftreten. Verfahren, die eine Anfälligkeit für diese Fehlerart aufweisen, haben eine verzerrte räumliche Wiedergabe zur Folge.

Die beschriebenen Effekte treten verstärkt auf wenn das Audiomaterial mehrfach hintereinander durch verlustbehaftete Verfahren kodiert wird. Nach Möglichkeit sollte dies daher vermieden werden.<sup>11</sup>

---

11 Vgl. Lerch (2008) S. 864 ff

## 4. Grundlagen der Streamingtechnologie

### 4.1 Das Client-Server Modell

Streamingangebote beruhen in der Regel auf dem Client-Server Modell, bei dem ein Server Daten zur Verfügung stellt und diese auf Anfrage des Clients in Paketen an selbigen über ein Netzwerk (z. B. das Internet) weitergibt. Für das Livestreaming von Radioprogrammen werden die Audiodaten dabei erst vom Encoder in ein streamingfähiges Format wie MP3, Windows Media Audio oder Ogg Vorbis gewandelt und mittels eines Streaming Protokolls an den Client weitergegeben. Dieser puffert die eingehenden Daten bis zu einem gewissen Maß, bevor die Wiedergabe auf der Empfängerseite starten kann.



Abb. 2 Prinzipieller Aufbau des Client Server Modells bei Streaming

Alle an diesem Vorgang beteiligten Komponenten können den Ablauf in der Übertragungskette beeinflussen. Somit hängt die Performance eines Streamingangebots von einer Vielzahl von Faktoren ab.<sup>12</sup>

### 4.2 Anforderungen an Streamingsysteme

Beim Design von Streamingsystemen sind einige Dinge zu bedenken und gewisse Anforderungen zu erfüllen, um einen reibungslosen Ablauf sicherstellen zu können:

Die Kontinuität des abgespielten Materials sollte gewährleistet sein und ab dem Zeitpunkt, an dem die Wiedergabe begonnen hat, möglichst ohne Unterbrechung vorstattengehen. Es gilt dabei einen Mittelweg zu finden zwischen ausreichender Pufferzeit vor dem Start und möglichst kurzer Verzögerung bis zum Anfang der Wiedergabe. Die Verzögerung setzt sich zusammen aus der Zeit, die es bedarf, um das Ausgangsmaterial zu enkodieren und die für die Übertragung erforderlichen Datenpakete zu erstellen sowie der Zeit, die für den Transfer vom Server zum Client benötigt wird und schließlich der Dauer des Pufferungsvorgangs bis zum Dekodieren. Diese Zeiten können schwanken und die Stetigkeit der Wiedergabe beeinträchtigen. Sollte das Netzwerk z. B. keine Quality-of-Service (QoS) garantieren können, was im Internet selten der Fall ist, so sind Latenzzeiten schlecht vorherzusagen. Auch der Deko-

<sup>12</sup> Vgl. Lee (2005) S. 3 ff

dierer kann aufgrund von parallel laufenden Prozessen auf der Clientseite unterschiedliche Verarbeitungszeiten aufweisen.

Weiter sollte natürlich eine (kosten-)effiziente Lösung angestrebt werden. Da Übertragungsbandbreite auch immer mit Kosten verbunden ist, muss man entscheiden zu welchen Bitraten ein Material zur Verfügung gestellt wird. Bei der Festlegung spielen dann wieder die Qualität verfügbarer Formate und Encoder eine Rolle sowie die Anzahl der erwarteten Nutzer. Falls viele Clients auf den Content zugreifen wollen, könnte es zu einer Überlastung und dasdurh einer Verschlechterung des Services des Systems kommen. Diese Problematik sollte bereits bei der Anschaffung der Hardwarekapazitäten berücksichtigt werden, indem ihre Dimensionierung sinnvoll gewählt wird.

Nicht zuletzt muss das System auch verlässlich sein. Hard- und Softwarefehler könnten den Server außer Gefecht setzen. Lösungen wie eine redundante Serverarchitektur müssen in Betracht gezogen werden. Diese ist natürlich wiederum mit finanziellem Aufwand verbunden, könnte aber auch im Falle von Überlastungen genutzt werden.<sup>13</sup>

### **4.3 Streamingfähige Protokolle**

Wie bereits erwähnt kann die Übertragungszeit Schwankungen unterzogen sein und die Funktionalität des Streamingsystems beeinflussen. Entscheidend dafür ist, wie die Transportprotokolle (Transportschicht im OSI Schichtenmodell) genutzt werden.

Für die Übertragung über das Internet kann man auf die Protokolltypen TCP (Transmission Control Protocol) und UDP (User Datagram Protocol) zurückgreifen. Die meisten Internet Applikationen wie WWW und FTP verwenden hierfür TCP. Es ist verbindungsorientiert und hält einen Datenstrom solange aufrecht, bis dieser explizit beendet wird. Ein Client kann gleichzeitig mehrere TCP basierende Verbindungen unterhalten, deshalb muss jede eindeutig identifizierbar sein. Zu diesem Zweck kommuniziert der Client mit dem Server vor der eigentlichen Datenübertragung und vereinbart eine Nummer, anhand der die Verbindung eindeutig zuzuordnen ist. Die Datenpakete erhalten ebenfalls eine Nummer, die Aufschluss darüber gibt, in welcher Reihenfolge sie beim Client verarbeitet werden müssen. Das Eintreffen von Gruppen von Datenpaketen wird an den Server zurückgemeldet, falls diese Bestätigung nicht eintrifft, werden die Pakete erneut versandt.

Das zweite Transportprotokoll UDP verfügt nicht über diese Kontrollmechanismen, es ist ein verbindungsloses Protokoll, bei dem die Datenpakete unabhängig voneinander transportiert

---

13 Vgl. Lee (2005) S. 11 ff

werden. Es werden keine Datenströme unterhalten, weshalb es einfacher aufgebaut sein kann. Die ganze Bandbreite kann für die Übertragung der Mediendaten genutzt werden, Kontrollfunktionen werden von der nächsten Schicht im OSI Schichtenmodell übernommen. In der Anwendungsschicht kann HTTP (Hypertext Transfer Protocol) verwendet werden, das eigentlich für die Übertragung von Webseiten entwickelt wurde und auf TCP aufbaut. Es hat viele Vorteile, beispielsweise dass für das Streaming normale Webserver anstatt spezieller Streamingserver genutzt werden können. Des Weiteren kommt es dadurch, dass die Daten wie beim herkömmlichen Webtraffic behandelt werden, zu weniger Konflikten mit Firewalls. Ausserdem sind in großem Maß Wiedergabeapplikationen kompatibel, da diese Art des Streamings von vielen Playern unterstützt wird. Die Nachteile der Verwendung dieses Protokolls liegen darin, dass die Übertragung aufwändig ist, denn die Kontrollfunktionen belegen Bandbreite und können den Datenstrom verlangsamen. Darüber hinaus muss sich der Stream die Bandbreite mit dem normalen Webtraffic teilen, was ebenfalls eine fehlerfreie Anlieferung der Daten behindern kann. HTTP wird auch vom Bayerischen Rundfunk für den MP3 Stream verwendet.

Daneben gibt es jedoch speziell entwickelte Streamingprotokolle, die die Ressourcen effizienter nutzen, denn das Streamen von Mediendaten stellt andere Anforderungen an den Datentransport als konventionelle Internetinhalte. Diese Streamingprotokolle können proprietär sein wie MMS von Microsoft (wird beim Windows Media Stream des BR eingesetzt) und RDT von RealNetworks, oder aber offene Standards wie RTSP (Real-Time Streaming Protocol) und RTP (Real-Time Transport Protocol). RTSP operiert von der Anwendungsschicht aus und wurde für die Steuerung kontinuierlicher Datenströme ausgelegt. Dabei können auch mehrere Streams gleichzeitig gesteuert werden, weshalb es auch als „Netzwerk Fernbedienung“ bezeichnet wird, denn über dieses Protokoll selbst werden keine Video- oder Audiodaten übertragen. Da RTSP unabhängig vom verwendeten Protokoll der darunterliegenden Transportschicht ist, kann es sowohl auf TCP wie auch auf UDP aufbauen. RTP wurde für den Transport von Daten für Echtzeitanwendungen entwickelt. Es generiert die Datenpakete, die dann häufig über UDP versendet werden. Als Kontrollprotokoll fungiert dabei das RTCP (Real-Time Control Protocol). RTP/RTSP optimieren somit die Datenübertragung über UDP oder TCP und stellen den Rahmen für die Implementierung von Quality-of-Service Mechanismen zur Verfügung.

In immer größerem Maße wird die Kombination RTSP/RTP/RTCP auch von Streamingservern mit proprietären Protokoll zugelassen, so dass sich diese Open Source Technologie als plattformübergreifendes Verfahren etabliert hat.<sup>14</sup>

---

14 Vgl. Lee (2005) S. 111 ff

## **5. Erläuterung der MUSHRA Methode**

Die ITU hat eine Empfehlung für den Vergleich von Mehrkanalkodierverfahren formuliert. In der Rec. ITU-R BS.1534-1 wird die sogenannte MUSHRA Methodology (MULTi Stimulus test with Hidden Reference and Anchor) beschrieben, die bereits mehrfach angewendet wurde und als verlässlich und aussagekräftig eingestuft werden darf.

### **5.1 Einsatzgebiet**

Die MUSHRA Methode wird eingesetzt für die subjektive Beurteilung von Audiomaterial, das im „mittleren“ Qualitätsbereich anzusiedeln ist, also deutlich wahrnehmbare Veränderungen durch die Umrechnung in ein anderes Format aufweist. Da Bitratenreduktion häufig unumgänglich ist und entsprechende Kodierverfahren eine Minderung der Qualität verglichen mit dem Ausgangsmaterial zur Folge haben, wurde darauf hingearbeitet, möglichst genau zu erfassen wie die Abweichung vom Original einzuordnen ist.<sup>15</sup> Die Vielzahl der gerade auch im Internet eingesetzten möglichen Formate wird bewertet und kann anschließend miteinander verglichen werden. Eine Skala von 0 (schlechte Qualität da stark verfremdet) bis 100 (dem Original entsprechend) liegt dieser Vorgehensweise zu Grunde und die gesammelten Ergebnisse können statistisch ausgewertet und interpretiert werden.

### **5.2 Voraussetzungen für die Durchführung**

Um möglichst eindeutige Ergebnisse erzielen zu können sollten einige formelle Voraussetzungen erfüllt sein, denn bewusst oder unbewusst herbeigeführte Effekte könnten das Ergebnis verfälschen.

#### **5.2.1 Reihenfolge der Materialien**

Die Reihenfolge, in der den Testpersonen die einzelnen Materialien zur Bewertung vorgeführt werden, sollte variieren. Auch innerhalb der verschiedentlich encodierten Materialien muss sich die Abfolge der angewendeten Kodierverfahren ändern. Dies dient einerseits dazu, Absprachen unter den Probanden zu unterbinden und andererseits um zu verhindern, dass sich ein erkennbares Muster herausbildet, das eine bestimmte Erwartungshaltung erwecken könnte, wo beispielsweise das dritte Kodierverfahren generell schlechter bewertet wird als das vierte.

---

15 Vgl. ITU-R BS.1534-1 S. 3

Auch ist die Qualitätsminderung in den verschiedenen Audioausschnitten nicht gleich einfach oder gleich schwierig zu erkennen. Durch eine zufällige Abfolge werden auch hier unbeabsichtigte Verfälschungen vermieden.<sup>16</sup>

### 5.2.2 Einführende Trainingssession

Eine Trainingssession mit wenigen Materialien, die vor dem eigentlichen Test durchgeführt werden sollte, dient dazu, die Testpersonen in den Umgang mit der Testsoftware einzuführen und sie mit der Art der zu erwartenden Veränderungen im Audiomaterial vertraut zu machen. Die Ergebnisse dieser Testsession dürfen allerdings nicht in die Bewertung des eigentlichen Tests mit einfließen.<sup>17</sup>

### 5.2.3 Abhörbedingungen

Natürlich sollten sich die Abhörbedingungen während des Tests nicht verändern, d. h. die Tests sollten möglichst im gleichen Raum mit der gleichen Technik durchgeführt werden. Die Aufstellung der Lautsprecher muss für die Bewertung von Mehrkanalton der Empfehlung der ITU-R BS.775 entsprechen. Auch die Lautstärke sollte, nachdem sie von der Testperson in der Trainingsphase festgelegt wurde, später nicht mehr verändert werden.

### 5.2.4 Auswahl des Audiomaterials

Die Aussagekraft eines MUSHRA-Tests hängt natürlich stark von der Auswahl des eingesetzten Audiomaterials ab. In der ITU Empfehlung zur Durchführung wird sogenanntes „kritisches Material“ verlangt, also einerseits solches, das typischerweise im Programm des Senders vorkommt, für den der Test gedacht ist. Andererseits sollen die Kodiersysteme auch an ihre Grenzen gebracht werden, um Probleme aufzuzeigen und deren Einfluss auf die Audioqualität bewerten zu können.

Innerhalb der Materialien sollen keine starken Lautheitsunterschiede auftreten, die Einfluss auf die Bewertung nehmen könnten.

Die Länge sollte bei etwa 10 - 20 Sekunden liegen.<sup>18</sup>

---

16 Vgl. ITU-R BS.1534-1 S. 3

17 Vgl. ITU-R BS.1534-1 S. 6

18 Vgl. ITU-R BS.1534-1 S. 5f

### 5.2.5 Auswahl der Testpersonen

Als Probanden kommen Menschen in Frage, die über normale Hörfähigkeiten verfügen (ISO Standard 389 wird als Richtlinie empfohlen), kritisches Hören gewohnt und somit auch fähig sind, in der beschriebenen Weise eine subjektive Wertung zur allgemeinen Audioqualität abgeben zu können.<sup>19</sup>

Um ein aussagekräftiges Ergebnis zu erhalten und entsprechende Schlüsse aus den Tests ziehen zu können, reichen 20 Testpersonen aus.<sup>20</sup>

### 5.2.6 Ausschlusskriterien in den gesammelten Bewertungen

Vor der Auswertung sollten die Daten auf extreme Evaluationen hin überprüft werden. Sollte eine Testperson überwiegend im oberen Bereich (excellent) oder im unteren Bereich (bad) bewertet haben, so könnte dies als zu unkritisch bzw. zu kritisch gelten und von der statistischen Analyse ausgeschlossen werden. Um das Ergebnis durch derartige Ausreißer im Bewertungsverhalten nicht zu beeinflussen, sollte erwogen werden, diese sich von der Mehrheit klar absetzenden Versuchspersonen nicht zu berücksichtigen. Weitere Kriterien können sinnvoll sein und das Ergebnis eindeutiger machen. So ist die Bewertung der Hidden Reference (das versteckte Original, das sich unter den verschieden kodierten Beispielen befindet) und des Anchors (ein bei 3,5 kHz tiefpassgefiltertes und deshalb als deutlich qualitätsgemindertes Signal anzusehen) ein Indiz dafür, ob der Test „richtig“ durchgeführt wurde.<sup>21</sup>

### 5.2.7 Statistische Analyse

Nachdem die Ergebnisse überprüft und die von der Auswertung auszuschließenden Evaluierungen eliminiert wurden, werden die Resultate statistisch aufbereitet und grafisch dargestellt. Als ersten Schritt errechnet man die arithmetischen Mittelwerte nach folgender Gleichung:

$$\bar{u}_{jk} = \frac{1}{N} \sum_{i=1}^N u_{ijk}$$

wobei  $u_i$  die Bewertung der Testperson  $i$  bei einem Kodierverfahren  $j$  und einem Audiobeispiel  $k$  ist

sowie  $N$  die Anzahl aller Testpersonen darstellt

---

19 Vgl. ITU-R BS.1534-1 S. 4

20 Vgl. ITU-R BS.1534-1 S. 5

21 Vgl. ITU-R BS.1534-1 S. 4f

Entsprechend können auch die Mittelwerte  $\bar{u}_i$  für jedes Kodierverfahren und  $\bar{u}_k$  für jedes Hörbeispiel berechnet werden.

Des Weiteren wird für jeden Mittelwert ein Konfidenzintervall (auch Vertrauensintervall) angegeben, das von der Standardabweichung und der Größe der Stichprobe abgeleitet wird und einen Bereich bezeichnet, in dem bei einem bestimmten Signifikanzniveau der tatsächliche Mittelwert liegt. Von der ITU wird ein Signifikanzniveau von 95% empfohlen, d. h. dass man davon ausgehen kann, dass der eigentliche Mittelwert zu 95% in dem, vom Konfidenzintervall beschriebenen Bereich liegt (dabei unterstellt man dass die Verteilung der einzelnen Bewertungen bestimmten Kriterien genügt). Die Konfidenzintervalle sind ein geeignetes Hilfsmittel, um Resultate beurteilen und miteinander zu vergleichen. Stellt man zwei Ergebnisse gegenüber, deren Konfidenzintervalle sich überlappen, kann man darauf schließen, dass sich diese nicht signifikant voneinander unterscheiden. Dementsprechend spricht man von einem signifikanten Unterschied, wenn sich die Vertrauensintervalle nicht überschneiden.

Die Berechnung erfolgt nach der Vorschrift:

$$\left[ \bar{u}_{jk} - \delta_{jk}, \bar{u}_{jk} + \delta_{jk} \right]$$

bei

$$\delta_{jk} = t_{0,05} \frac{S_{jk}}{\sqrt{N}}$$

Der t-Wert bezeichnet dabei das Signifikanzniveau,  $t_{0,05}$  steht hier für den gewählten Wert von 95%.

Die Standardabweichung wird folgendermaßen ermittelt:

$$S_{jk} = \sqrt{\sum_{i=1}^N \frac{(\bar{u}_{jk} - u_{ijk})^2}{(N-1)}}$$

Zur besseren Veranschaulichung der Ergebnisse wird die grafische Darstellung für die einzelnen Testsequenzen empfohlen, anstatt nur die Mittelwerte und Vertrauensintervalle der getesteten Audiomaterialien zusammengefasst darzulegen.<sup>22</sup>

Bei der Auswertung der gesammelten Daten wird bei einem MUSHRA Test davon ausgegangen, dass die Bewertungen normalverteilt sind. Die Auswertung erfolgt somit nach den Gesichtspunkten eines parametrischen Tests, d. h. dass angenommen wird, dass die Verteilung der Testergebnisse bekannt und für alle Tests gleich ist. Obwohl dies nicht überprüft wird und

<sup>22</sup> Vgl. ITU-R BS.1534-1 S. 12f



somit auch von dieser Annahme abweichen kann, hat sich dieses Vorgehen bei MUSHRA-Tests etabliert, die Ergebnisse werden als aussagekräftig angesehen.

### 5.2.8 Der Wilcoxon-Vorzeichen-Rang-Test als ergänzendes Mittel zur statistischen Analyse

Will man zwei Fälle näher untersuchen, bedient man sich des Wilcoxon-Vorzeichen-Rang-Tests. Dieser darf auch dann angewendet werden, wenn man keine Angaben über die Verteilung hat. Die Daten werden mit diesem Test effektiver genutzt und sind somit für den Vergleich von paarweise durchgeführten Stichproben sehr gut geeignet.

Beim Wilcoxon-Vorzeichen-Rang-Test werden folgende Schritte nacheinander durchgeführt:

1. Man bildet die Differenz für jeden Wert der untersuchten Stichproben

$$D_i = X_i - Y_i \quad i = 1, \dots, n$$

wobei  $X_i$  Stichprobe eins und  $Y_i$  Stichprobe zwei darstellt

2. Man bildet die dazugehörigen betragsmäßigen Differenzen  $|D_1|, \dots, |D_n|$

3. Die betragsmäßigen Differenzen werden der Größe nach geordnet und Rängen zugeteilt, wobei der kleinste Betrag den Rang 1, der zweitkleinste Betrag den Rang 2 usw. erhält.

Nun bildet man die Summe aus den Rängen nach der Vorschrift

$$W^+ = \sum_{i=1}^n \text{rg}|D_i| Z_i \quad \text{mit} \quad Z_i = \begin{cases} 1 & \text{wenn } D_i > 0 \\ 0 & \text{wenn } D_i < 0 \end{cases}$$

Dabei bezeichnet  $\text{rg}|D_i|$  den Rang von  $|D_i|$ , die Variable  $Z_i$  hat nur eine Auswahlfunktion, so dass nur die Ränge für positive  $D_i$  vor der Betragsbildung aufsummiert werden. Anhand einer Tabelle für einen Wilcoxon-Vorzeichen-Rang-Test kann nun entschieden werden, ob sich die Stichproben bei einem bestimmten Signifikanzniveau unterscheiden.<sup>23</sup>

23 Vgl. Fahrmeir, Künstler, Pigeot, Tutz (2001) S. 428 f

## 6. Durchführung der Hörtests

### 6.1 Auswahl der Audiobeispiele

Als Hörbeispiele sollten Tonproduktionen in den Test aufgenommen werden, die so auch in den Programmen Bayern 2 und BR Klassik wieder zu finden sein könnten. Bei der Zusammenstellung hatte ich mich für Ausschnitte aus folgenden Gattungen entschieden: Hörspiel, Weltmusik, Jazz, Elektronische Popmusik, Oper, Operette, Klassisches Streichquartett, Klassische Musik mit großem Orchester, Barocke Musik, Orchesterale Filmmusik.

Aus den Ergebnissen früher durchgeführter MUSHRA Tests zum Vergleich von Mehrkanalkodierverfahren (z. B. EBU Tech 3324 - Evaluation of Multichannel Audio Codecs) war mir bekannt, dass es Signaltypen gibt, bei denen Artefakte besonders deutlich auftreten, z. B. Applaus, Cembalo oder Mundharmonika. Es war mein Bestreben, dass diese schwierigen Materialien ebenfalls im Test enthalten sind. So hatte ich ein Barockstück mit Cembalo und ein Weltmusikstück mit Mundharmonika ausgewählt. Der Applaus war als eigenständiges Hörbeispiel Teil des Tests.

Bei der Auswahl hatte ich mich zudem dazu entschlossen, dass ich auf unterschiedliche Tonträger zurückgreife (SACD, DVD-Audio, DVD-Video), um Ausgangsmaterial in verschiedenen Formaten einbringen zu können.

Folgende Tabelle gibt die Herkunft und das Format der Audioausschnitte an:

Titel	Gattung	Quelle	Format
Maná - Vivir sin aire	Weltmusik	DVD-Audio	MLP 5.1, 24bit, 96kHz
Gustav Mahler - Allegro Maestoso aus 2. Sinfonie	Klassische Musik (großes Orchester)	DVD-Audio	MLP 5.1, 24bit, 96kHz
Crystal Method - Starting Over	Elektronische Popmusik	DVD-Audio	LPCM 5.1, 24bit, 48kHz
Manfred Leuchter - Ma Fi Minnak	Jazz	DVD-Video	DTS 5.0, 1536kbps
Jules Verne - Die Reise zum Mittelpunkt der Erde	Hörspiel	DVD-Video	DTS 5.0, 768kbps
Pjotr Iljitsch Tschaikowski - 4. Sinfonie	Klassische Musik (großes Orchester)	DVD-Video	DTS 5.0, 1536kbps
Johann Strauß - Wiener Blut	Operette (Orchester und Gesang)	DVD-Video	Dolby Digital 5.1, 448kbps
Giuseppe Verdi - Gefangenchor aus Nabucco	Oper (Orchester und Chor)	DVD-Video	Dolby Digital 5.1, 448kbps

Wolfgang Amadeus Mozart - Rondo Allegro aus Eine kleine Nachtmusik	Klassisches Streichquartett	DVD-Video	Dolby Digital 5.1, 448kbps
Giuseppe Verdi - Nabucco	Applaus	DVD-Video	Dolby Digital 5.1, 448kbps
Marlene Streeruwitz - Wunschzeit	Hörspiel	SACD	DSD
Johann Christian Bach - Andante e grazioso aus Konzert in f-moll	Barockmusik	SACD	DSD
Francis Lai - Love Theme aus Love story	Orchestrale Filmmusik	SACD	DSD

In die Trainingssession hatte ich auf vier Materialien zurückgegriffen, die auch in der engeren Wahl für den eigentlichen Test gestanden waren:

Titel	Genre	Quelle	Format
Johann Sebastian Bach - Adagio aus BWV 1060	Barockmusik	DVD - Audio	MLP 5.1, 24bit, 96kHz
Adriana Hölszky - Gemälde eines Erschlagenen	Vokalmusik	DVD-Video	DTS 5.1, 1536kbps
Neil Young & The Band - Helpless	Popmusik	DVD-Video	Dolby Digital 5.1, 448kbps
Maurice Jarre - Theme aus Lawrence von Arabien	Orchestrale Filmmusik	SACD	DSD

## 6.2 Formate im Vergleich

### 6.2.1 MP3 Surround

Das wohl gängigste Audiokodierverfahren ist das vom Fraunhofer Institut entwickelte MP3 (MPEG-1 Audio Layer 3) Format. Es findet zahlreiche Verwendung in vielen Bereichen, z. B. bei der Verbreitung von Musik über das Internet. Für die Verarbeitung von Mehrkanalton setzt das wahrnehmungsangepasste Verfahren die Technik des Binaural Cue Coding ein, wie es in 3.2.3 beschrieben ist. Das als MP3 Surround bezeichnete Verfahren wird zum Teil auch schon für das Internetangebot von Radiostationen, auch in Bayern, genutzt, z. B. Rockantenne und Antenne Bayern.

### 6.2.2 Windows Media Audio

Das proprietäre Kodierverfahren von Microsoft ist mit dem Codec Windows Media Audio 10

Professional auch für Mehrkanalanwendungen geeignet. Für das Internetstreaming werden die .wma-Dateien in das ebenfalls von Microsoft entwickelte Containerformat mit der Endung .asf angeboten.

### 6.2.3 Ogg Vorbis

Die Stiftung Xiph.org entwickelte das patent- und lizenzfreie Containerformat Ogg und den Audiocodec Vorbis als Alternative zu den MPEG Technologien und gilt als effektivere Lösung im Vergleich zu MP3. Dieses Open Source Format verwendet eine MDCT mit adaptiven Fensterlängen, Entropiekodierung und Vektorquantisierung bei der Berechnung der Mithörschwellen. Bei der Mehrkanalkodierung wird die Datenreduktion an die Bandbreite des LFE-Kanals angepasst. Die Dateiendung lautet .ogg.<sup>24</sup> Ogg Vorbis ist vor allem für Streaming ausgerichtet und wird z. B. bei Deutschlandradio Kultur und Deutschlandfunk für das Streaming des Programms (in Stereo) mit variabler Bitrate verwendet.

### 6.2.4 Dolby Digital

Das Verfahren der Firma Dolby Laboratories arbeitet bei der Segmentierung und Transformation für die psychoakustische Kodierung mit einer adaptiven MDCT, in jedem Kanal wird die Fensterlänge unabhängig zu den anderen Kanälen adaptiert. Eine Entropiekodierung kommt nicht zum Einsatz, allerdings wird bei hoher Informationsdichte eine frequenzselektive MS-Matrizierung angewendet um Redundanzen zwischen den Kanälen zu verringern.<sup>25</sup> Der LFE Kanal wird steilflankig bei 120 Hz bandbegrenzt. Die Dateiendung von Dolby Digital lautet .ac3, diese wurde vom Bitstream-Format Adaptive Transform Coder 3 abgeleitet.<sup>26</sup> Für die digitale Satelitenausstrahlung wird das Verfahren bei einer Datenrate von 448 kbps bereits vom Bayerischen Rundfunk verwendet und kann im Test als Referenz für eine bereits eingesetzte Technik herangezogen werden.



24 Vgl. Görne (2008) S. 193

25 Vgl. Görne (2008) S. 190

26 Vgl. Slavik und Weinzierl (2008) S. 627

### 6.3 Eingesetzte Software bei der Vorbereitung der Hörtests

Nachdem ich mich umfassend mit dem mir zur Verfügung stehenden mehrkanaligen Audio-material beschäftigt und mich entsprechend eingehört hatte, musste ich die verschiedenen Formate in eine einheitliche Form bringen, die von den Encodern akzeptiert werden würde. So erstellte ich aus den jeweiligen Ausgangsmischungen für jeden Kanal eine linear PCM-kodierte Datei im WAV-Format mit 24 bit und 48kHz.

Für das Material von DVD-Audio verwendete ich die freie Software **DVD-Audio Explorer 2008 (Alpha 8)**, die einerseits aus dem Bitstream der DVD die Audiospur extrahierte und andererseits aus den MLP-kodierten Daten WAV Dateien erzeugte. Beim anschließenden Import in die Tonbearbeitungssoftware **Nuendo (Version 3.2.0)** wurden die Dateien, die ursprünglich mit 96 kHz abgetastet worden waren, auf eine Samplingfrequenz von 48 kHz heruntergerechnet.

DTS oder Dolby Digital kodierte Audiodaten von DVD-Video konnte ich mit dem Programm **DVD Audio Extractor (Version 4.5.4)** dem DVD-Bitstream entnehmen und in das gewählte WAV-Format umwandeln.

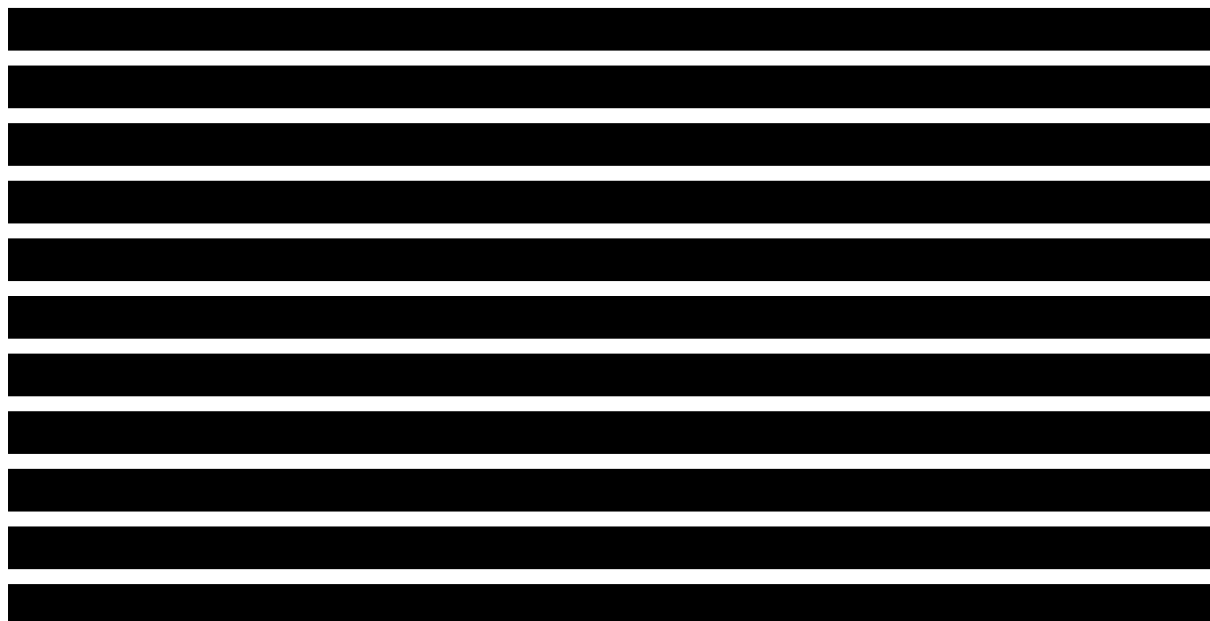
Da es leider nicht möglich ist, auf DSD-Daten von einer SACD rechnerintern zuzugreifen und weiter zu verarbeiten, musste ich von diesem Medium über die Analogausgänge des Players in die Lineeingänge meines Audiointerfaces, einem Focusrite Saffire Pro 26, nach Nuendo überspielen.

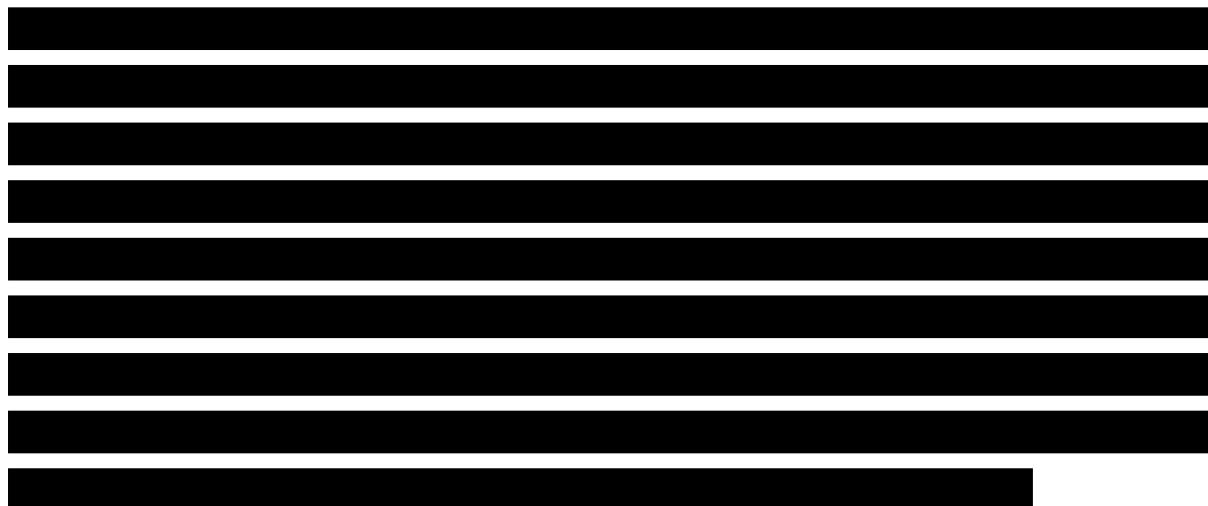
Nachdem ich mein Ausgangsmaterial im einheitlichen Format in Nuendo importiert hatte, ging ich daran, die Einzelkanäle in der Reihenfolge L, R, C, LFE, SL, SR übereinander zu ordnen, sodann die Tonbeispiele in ihrer Länge festzulegen, entsprechend auszuschneiden und, nachdem ich die Lautheit der Hörbeispiele untereinander angepasst hatte, als sechskanaliges Interleavedfile (wiederum 24 bit, 48 KHz) zu exportieren.

Nun konnte ich den Encodiervorgang starten. Um die Windows Media Dateien zu erzeugen, griff ich auf die Software **Microsoft Windows Media Encoder 9 Series (Version 9.00.00.2980)** zurück, welche auf der Microsoft Homepage kostenlos zur Verfügung steht. Ich verwendete dabei die von der Software angebotenen konstanten Datenraten 128 kbps, 192 kbps und 384 kbps; der der Software zugrundeliegende Codec war dabei Windows Media Audio 10 Professional mit der Einstellung Two-pass encoding. Die daraus resultierenden .wma-Dateien importierte ich wiederum in Nuendo, wo sie automatisch in sechskanalige interleaved WAV-Dateien umgewandelt wurden und für den Test eingesetzt werden konnten.

Für die Erstellung der MP3 Surround Dateien steht auf der Seite des Fraunhofer Institutes für

Integrierte Schaltungen ein Encoder kostenlos zum Download bereit, in meinem Fall war es der **MP3 Surround Gui-Encoder V1.1.5** mit der **MP3 Surround Library V04.01.01**. Die Software ist auf eine konstante Bitrate von 192 kbps eingestellt, eigene Einstellungen können nicht vorgenommen werden. Um das Material wieder zu dekodieren musste ich es mit der Abspielsoftware **Winamp v5.56** (in Winamp ist ab der Version 5.5 ein MP3 Surround Decoder standardmäßig integriert) und sechs analoge Kanäle eines M-Audio FW 410 Firewire-interfaces wiedergeben, um es anschließend wiederum über das Focusrite Interface in Nuendo aufzunehmen. Dabei stellte ich die Eingangsvorverstärker mit einem zuvor angelegten Sinuston so ein, dass alle Kanäle mit dem gleichen Pegel angezeigt wurden und die Überspielung eins zu eins erfolgen konnte. In Nuendo musste ich nun die neu entstandenen WAV-Dateien so bearbeiten, dass sie in Position und Länge zum Ausgangsmaterial passten, bevor ich sie anschließend wieder in das angestrebte Format (wav interleaved 24 bit 48 kHz) exportierte. Den Enkodiervorgang für Ogg Vorbis wickelte ich über die freie Software **oggdropXPd V1.9.0 (libvorbis ao TuVb5.7)** ab. Bei der Einstellung Standard Quality Mode unter dem Punkt General Encoder Options wählte ich ein Quality Setting von -1,00, um Dateien mit variabler Bitrate zu erhalten, deren Größe etwas geringer war als diejenigen, die mit Windows Media und MP3 Surround Encoder bei einer Datenrate von 192 kbps erstellt wurden. Die Umwandlung der OGG- in WAV-Dateien erfolgte mit der gleichen Software. Für Dolby Digital verwendete ich die Encoding Applikation **Apple Compressor Version 3.0.5**. Die damit erzeugten sechskanaligen AC3-Dateien konnten durch den Import in Nuendo, wie schon bei Windows Media, in das entsprechende WAV-Format konvertiert werden.





#### **6.4 Auswahl der Testpersonen**

Als Testpersonen konnte ich Tontechnikerkolleginnen und -kollegen aus den Abteilungen Studioproduktion und Betrieb im Bereich Hörfunk des Bayerischen Rundfunks für meine Zwecke gewinnen. Sie stellen ideale Testhörer dar, da sie einerseits mit den Programmen Bayern 2 und BR Klassik vertraut sind und andererseits kritisches Hören zu ihren täglichen Aufgaben im Berufsleben gehört. Dass die Probanden ein intaktes Gehör besitzen habe ich dabei vorausgesetzt, denn ansonsten könnten sie ihrer Beschäftigung nicht nachgehen. Die Bewertungen basieren somit auf einer fachlichen Kompetenz der Testhörer und können daher als durchaus substanzhaltig angesehen werden.

#### **6.5 Testsoftware**

Für die Durchführung der Tests wurde mir von Seiten des Bayerischen Rundfunks die Software STEP (Subjective Testing and Evaluation Program) der Firma Audio Research Labs in der Version 1.05 zur Verfügung gestellt. Um dem Programm mitzuteilen, welche Dateien verwendet werden, welches davon das Original ist und wann ein neues Beispiel beginnt, muss vor Beginn ein Textfile (sog. Session-File) nach einer bestimmten Struktur erstellt werden, das dann in dem Ordner, in dem sich die Audiodaten befinden, mit der Endung .asi abgespeichert wird. In der Startmaske von STEP lässt sich diese Datei öffnen, auf Fehler überprüfen und starten.

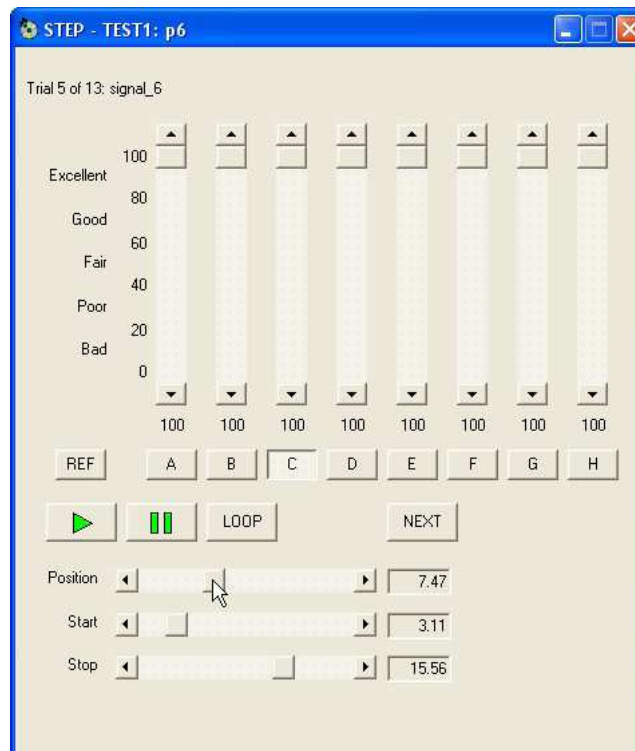


Abb. 3 Bedienoberfläche von STEP während eines Tests

Mit STEP ist ein direkter Vergleich der Hörbeispiele möglich. Ein unterbrechungsfreies Umschalten zwischen den parallel ablaufenden Audiodateien erlaubt die bestmögliche Gegenüberstellung der Unterschiede, wie sie von den einzelnen Encodern erzeugt werden. Über den Button REF kann man immer wieder mit dem unbearbeitete Original vergleichen, die Buttons A - F stehen für die sieben en- und dekodierten Dateien sowie das Hidden Reference und Anchor Signal. Über jedem dieser Buttons befindet sich eine Skala von 0 - 100, die weiter in fünf Stufen (Excellent - Good - Fair - Poor - Bad) im Abstand von 20 Punkten unterteilt ist, mit der anhand eines Schiebereglers die Qualität des Materials bewertet wird. Das Abspielen der Beispiele kann unterbrochen und geloopt werden, des Weiteren symbolisiert ein horizontal mitlaufender Balken die Position des ablaufenden Elements in der Zeitleiste. Die nebenstehende Anzeige in Sekunden und Millisekunden gibt Auskunft über den gerade abgespielten Zeitpunkt. Darüber hinaus lässt sich der Abschnitt, den man vergleichen möchte, eingrenzen. Anhand der beiden querliegenden Balken Start und Stop können Anfangs- und Endpunkt eingestellt werden. Ist man beispielsweise der Meinung, dass an einer ganz bestimmten Stelle die Unterschiede am deutlichsten erkennbar sind, kann diese eingegrenzt und geloopt werden, um dann zwischen dem Original und den einzelnen bearbeiteten Materialien hin- und herzuschalten und so seine Wertung zu finden.

Ist die Bewertung für ein Hörbeispiel abgeschlossen, drückt man die Schaltfläche NEXT und



gelangt zum nächsten, bzw. am Ende der Testreihe in die Startmaske. Einen Button um zurück zu gehen ist nicht vorgesehen. Die Zahlenwerte die in den Skalen über den Buchstaben eingestellt wurden, werden nach Abschluss der Testreihe neben den jeweiligen Dateinamen im Hintergrund automatisch in ein Textfile geschrieben, so dass hinterher eine eindeutige Zuordnung gewährleistet ist. Über eine Voreinstellung können die einzelnen Beispiele durchmischt und in zufälliger Reihenfolge wiedergegeben werden, ebenso werden innerhalb eines Beispiels die Schaltflächen mit den Buchstaben immer mit anderen Materialien belegt.

## 6.6 Testumgebung

### 6.6.1 Raum

Zur Durchführung der Hörtests wurde mir von Seiten des Bayerischen Rundfunks im Funkhaus München das Studio DSR5 (Digitaler Schneiderraum 5) zur Verfügung gestellt. Dieses wird normalerweise für die Bearbeitung von Musikaufnahmen von Tonmeistern und Tontechnikern genutzt. Es ist ein Raum mit einer Grundfläche von ca. 24 m<sup>2</sup> und einer mittleren Höhe von etwa 2,25 m.

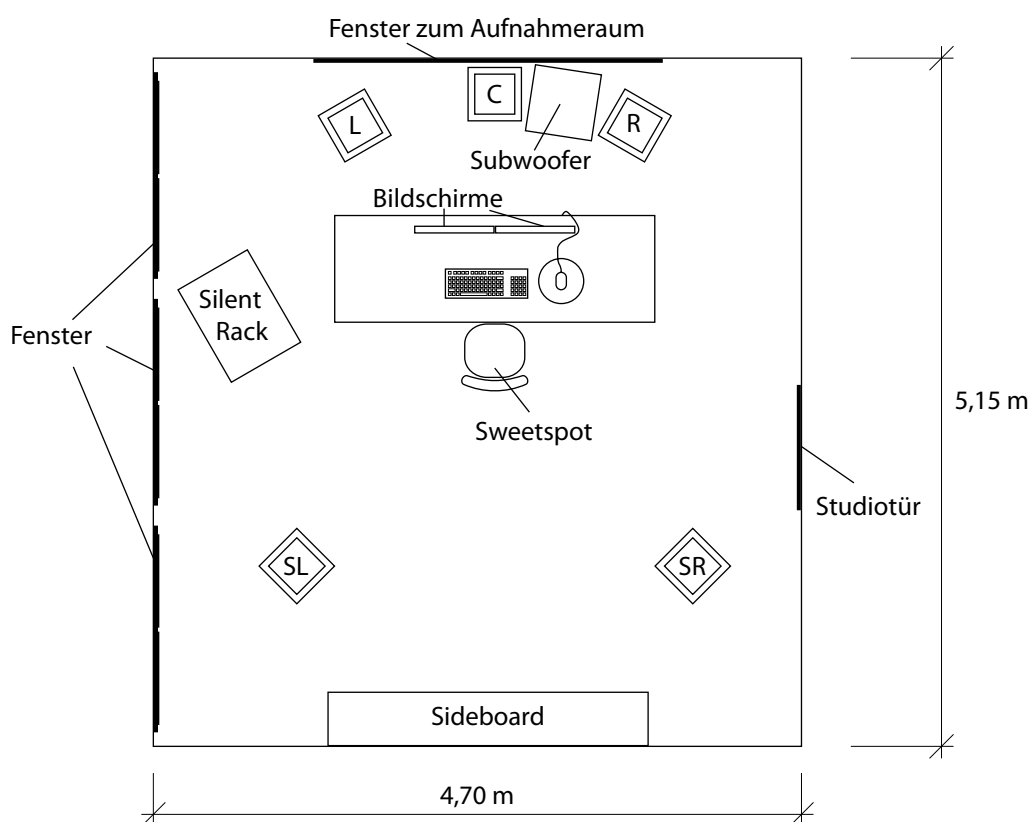


Abb. 4 Skizze Einrichtung DSR5

Alle Wände sind mit Absorbern ausgestattet, größtenteils mit geschlitzten Holzplatten, die Decke sowie die Tür mit Lochblechelementen zur Schalldämpfung. An der Fensterfront sind

Vorhänge zur Verbesserung der Akustik angebracht, ebenso am Fenster zum Aufnahme-  
raum.



Abb. 5 Blick auf Arbeitsplatz und Frontlautsprecher DSR5

### 6.6.2 Eingesetzte Technik

**Rechner:** Die STEP Software war auf meinem Notebook installiert, einem Fujitsu Siemens Amilo Xi 1526 mit Genuine Intel CPU bei 1,83 GHz und 1 GB RAM. Das Betriebssystem war Windows XP Home Edition Version 2002 mit Service Pack 3, v. 3244. Sämtliche Audiofiles wurden auf diesem Rechner bearbeitet und während des Tests von der lokalen Festplatte abgespielt.

**Audiointerface:** Während des Tests verwendete ich ein Fireface 400 der Firma RME mit der Firmware/Hardware Revision 1.63 und der Treiberversion 2.95.

**Lautsprechersystem:** Die fünf Fullrange Lautsprecher waren von der Firma ME Geithain jeweils das Regielautsprechermodell RL 903K. Als Subwoofer stand mir ebenfalls ein Produkt von ME Geithain zur Verfügung, das Modell Basis 3 Active Bass Loudspeaker.

Vom Interface aus wurden die Kanäle C, LFE, SL, SR direkt über Klinckenkabel mit den Lautsprechern verbunden. Die Kanäle L und R wurden über den Subwoofer geführt bevor sie auf den linken und rechten Frontlautsprecher geschickt wurden, um das Bassmanagement des Subwoofers für diese Kanäle zu nutzen. Die Einstellung wurde vor dem ersten Test einmal so vorgenommen, dass der Bassbereich am Sweetspot weder überbetont noch vernachlässigt wurde. Ebenso wurden vor dem ersten Test die Abstände und Winkel vom Sweetspot zu den Lautsprechern überprüft und die Pegel anhand von rosa Rauschen für alle Hauptkanäle gemessen.

## **6.7 Trainingssession**

Während der Trainingsession wurden den Probanden die Funktionen der Testsoftware erläutert und sie hatten die Möglichkeit sich anhand der vier Testhörbeispiele einzuhören und auf die Aufgabenstellung einzustellen. Das Audiointerface stellt einen Softwaxemixer zur Verfügung, mit dem die gewünschte Abhörlautstärke eingestellt werden konnte, die dann unverändert blieb. Die Testpersonen wurden von mir angewiesen, dass sie die allgemeine Audioqualität der einzelnen Kodierverfahren im Vergleich zum Original bewerten sollten, wobei die höchste Wertung das Signal erhalten sollte, das am wenigsten vom Original abweicht. Zudem wurde auf Hidden Reference Anchor hingewiesen und deren Bedeutung für den Test erklärt. Während der Trainingsession war ich im Raum um eventuell auftommende Fragen mit den Testhörern zu klären und anschließend die eigentliche Session zu starten. Anschließend habe ich den Raum verlassen, so dass die Probanden sich in Ruhe auf das Material einlassen konnten.

## **6.8 Durchführungszeitraum und Dauer der Tests**

Die Tests fanden in der Zeit vom 8. September bis zum 2. Oktober im DSR5 im Funkhaus des Bayerischen Rundfunks in München statt. Pro Test wurden etwa anderthalb bis drei Stunden benötigt. Da konzentriertes Hören über eine derart lange Zeit sehr anstrengend ist, war es den Probanden natürlich erlaubt, nach Bedarf Pausen einzulegen. Ein zeitliches Limit, indem die Testreihe abgeschlossen sein musste, wurde nicht gesetzt.

## 7. Auswertung und grafische Aufbereitung

Für die Auswertung der Ergebnisse verwendete ich Microsoft Office Excel 2003, wo alle Bewertungen aus den von der Testsoftware erstellten Textdateien importiert und anhand von Filtern entsprechend verarbeitet wurden. Für die statistischen Berechnungen und Erstellung der Diagramme arbeitete ich mit WinSTAT für Microsoft Excel (Version 2007.1), einem Statistik-Add-In für Microsoft Excel. Dieses Programm stellt Funktionen zur statistischen Analyse innerhalb von Excel zur Verfügung, somit können die in Excel organisierten Daten sofort weiter ausgewertet werden.

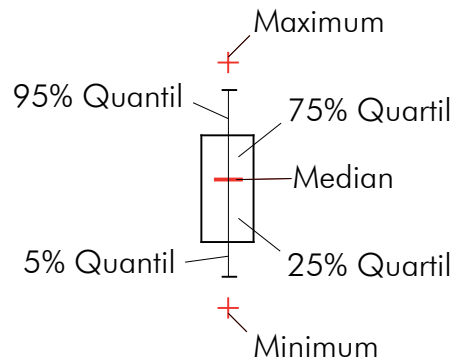
Die Grafiken sind jeweils so aufgebaut, dass links die Bewertungsskala von 0 bis 100 aufgetragen ist und rechts die Abstufungen in die einzelnen, 20 Bewertungspunkte umfassenden Bereiche Excellent (100 - 80), Good (79 - 60), Fair (59 - 40), Poor (39 - 20) und Bad (19 - 0). Die jeweiligen Diagramme sind dann nach Kodierverfahren oder Hörbeispielen gruppiert, die entsprechende Einteilung findet sich an der Unterseite der Diagramme.

Bevor ich die Arbeit in Excel begonnen habe, wurden sämtliche abgegebenen Bewertungen von mir nochmals überprüft und bereinigt. Insgesamt haben 20 Personen am Test teilgenommen, die 13 Hörbeispiele mit je neun Fällen evaluierten. Wurde die Hidden Reference nicht im Excellent Bereich eingestuft, somit nicht als das Original erkannt, habe ich alle Werte einer Testperson für das entsprechende Hörbeispiel ausgenommen. Ebenso, wenn der Anchor zu hoch, also mit 20 Punkten und mehr beurteilt wurde. Der dritte Fall war eine offensichtlich Fehlbedienung der Software, oder zu kritische bzw. unkritische Bewertung, beispielsweise für acht von neun Vergleichsmöglichkeiten zur Referenz 100 Punkte zu vergeben. Es wurden 37 mal neun Wertungen von mir eliminiert, so blieben am Schluss 223 mal neun Wertungen, also 2007 Zahlenwerte, die ausgewertet werden konnten.

### 7.1 Boxplot

Um einen ersten Eindruck darüber zu geben, wie die Bewertungen für die einzelnen Kodierverfahren ausgefallen sind und wie groß deren Streuung ist, möchte ich den Boxplot für jedes Kodierverfahren über alle Materialien der Auswertung nach der MUSHRA Methode voranstellen.

Ein Boxplot kann folgendermaßen interpretiert werden:



Der Median gibt den mittleren Wert in der Liste aller Bewertungen an und teilt diese in zwei gleiche Hälften, demzufolge sind die Hälfte aller Bewertungen größer oder gleich dem Median und ebenso die Hälfte aller Bewertungen kleiner oder gleich dem Median. Bei einer ungeraden Anzahl von Werten ist der Median genau der mittlere, bei einer geraden Anzahl von Werten wird er aus dem arithmetischen Mittel der beiden mittleren gebildet.

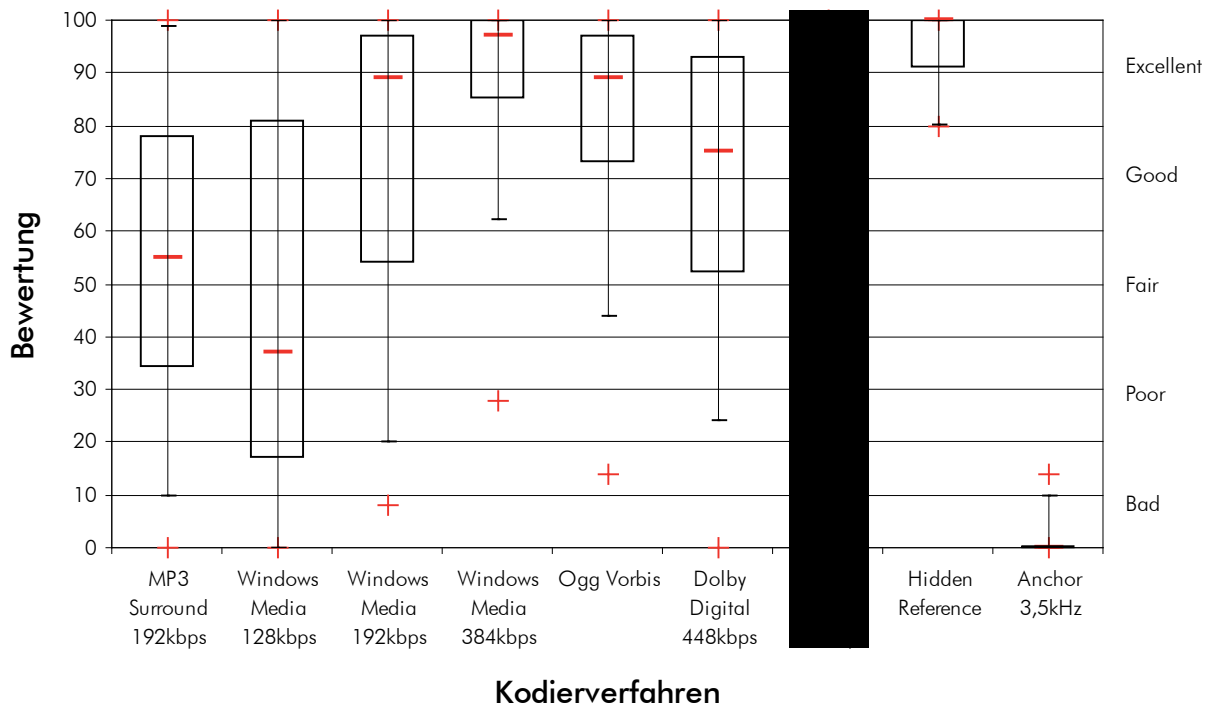
Das 75% Quartil und das 25% Quartil bilden die Box (hier durch den schwarzen Rahmen markiert), in der sich 50% aller Bewertungen befinden. Die Grenzen an der oberen und unteren Seite der Box geben Auskunft darüber in welchem Bereich sich ein Viertel aller Werte ober- und unterhalb des Medians befinden.

Das 95% Quantil und das 5% Quantil zeigen an, wo 90% aller Bewertungen zu finden sind, wobei die oberen und unteren 5% ausgespart werden.

Maximum und Minimum bezeichnen die höchste, bzw. die niedrigste Bewertung.<sup>27</sup>

---

27 Vgl. Fahrmeir, Künstler, Pigeot, Tutz (2001) S. 65 f



Die Tabelle gibt zusätzlich Informationen, in Zahlen ausgedrückt, über die Mittelwerte, den Median und die Standardabweichung der Bewertungen für die jeweiligen Kodierverfahren sowie für Hidden Reference und Anchor.

Kodierverfahren	MP3 Surround 192kbps	Windows Media 128kbps	Windows Media 192kbps	Windows Media 384kbps	Ogg Vorbis	Dolby Digital 448kbps		Hidden Reference	Anchor 3,5kHz
Median	55	37	89	97	89	75		100	0
Mittelwert	55,22	45,73	75,52	90,89	82,59	70,50		95,21	1,10
Std. Abweichung	27,22	33,29	27,06	13,16	18,23	25,06		6,40	2,85

Es können an dieser Stelle bereits viele Erkenntnisse gewonnen werden:

Bei nicht einmal der Hälfte aller Fälle erhält das **MP3 Surround** Kodierverfahren von den Testpersonen eine gute oder exzellente Bewertung. Die relativ hohe Standardabweichung weist zudem auf eine äußerst divergente Beurteilung des Hörempfindens bei dementsprechend eingestellter Datenrate hin. Zwar lassen sich mit dem Verfahren teils durchaus positive Ergebnisse erzielen - immerhin wurde in rund einem Viertel aller Fälle eine exzellente Bewertung ausgesprochen. Doch bei einigen speziellen Hörbeispielen vermag das MP3 Surround-Verfahren nur unbefriedigende Resultate zu liefern, wie in 7.2.2 noch zu sehen sein

wird. Überwiegend scheinen die Testpersonen aber deutliche Unterschiede zum Original ausgemacht zu haben.

Bei **Windows Media mit 128 kbps** befinden sich mehr als die Hälfte aller Bewertungen im Poor und Bad Bereich. Die Probanden scheinen somit deutliche Qualitätsminderungen festgestellt zu haben. Die Box reicht vom Bad- bis in den Excellent Bereich und die Standardabweichung liegt nochmals etwas höher als bei MP3 Surround, daher kann man davon ausgehen, dass Windows Media bei einer Bitrate von 128 kbps eine schlecht vorhersagbare Qualität liefert, was als durchaus problematisch angesehen werden müsste. Da der Mittelwert mehr als acht Punkte über dem Median liegt, ist zu vermuten, dass extreme Ausreißer nach oben hin den Mittelwert deutlich anheben. Dies zeigt: Auch wenn wie oben gesehen in etwa jede zweite Bewertung negativ ausfällt, beurteilen die Testhörer das Verfahren in einzelnen Fällen doch durchaus positiv. So lässt sich anhand der Boxplot-Grafik erkennen, dass etwa jede vierte Bewertung in den exzellenten Bereich fällt. Der komplette Bereich von 0 bis 100 wurde von den Probanden ausgenutzt, mindestens 5% aller Bewertungen wurden mit 0 und ebenso mindestens 5% mit 100 bewertet. Eine detaillierte Analyse des Kodierverfahrens nach Hörbeispielen folgt in 7.3.

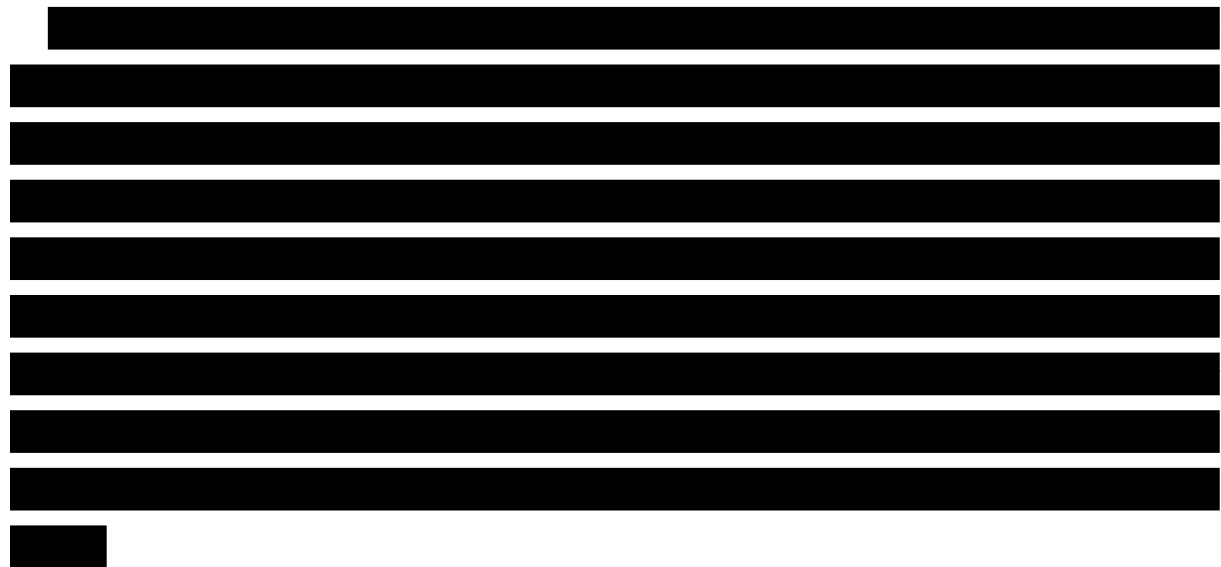
Für **Windows Media mit 192 kbps** ist festzustellen, dass mehr als die Hälfte aller Wertungen im Excellent Bereich angesiedelt sind, knapp 50% weisen sogar 90 Punkte und mehr auf. Der Median weist die deutlichste Abweichung in allen Fällen vom Mittelwert auf, jedoch wirken sich hier im Gegensatz zu Windows Media 128 kbps Ausreißer nach unten negativ auf den Mittelwert aus. Auffällig ist zudem, dass in mehr als drei von vier Fällen eine Wertung mit über 50 Punkten abgegeben wurde. Die Standardabweichung ist in etwa so groß wie bei MP3 Surround, die Streuung also relativ hoch. Dies wird in der weiteren Auswertung noch zu beachten zu sein, denn die Vermutung liegt nahe, dass nur für bestimmte Signalarten das Kodierverfahren an seine Grenzen stößt.

**Windows Media mit 384 kbps** hat den höchsten Median, den höchsten Mittelwert und die geringste Standardabweichung unter allen Kodierverfahren. 75% aller Bewertungen liegen im Excellent Bereich und 95% mindestens im Good Bereich. Einige Ausreißer nach unten wirken sich zwar negativ auf den Mittelwert aus, der aber trotzdem noch über 90 Punkten liegt, was darauf schließen lässt, dass es sich nur um einzelne extreme Wertungen handelt. Insgesamt scheint das Verfahren aber mit dieser Bitrate konstant gute Ergebnisse zu liefern.

**Ogg Vorbis** scheint mit über 50% aller Bewertungen im Excellent Bereich und mehr als Dreiviertel der Werte über 70 Punkten ebenfalls überwiegend gute und sehr gute Ergebnisse

bieten zu können. Dafür spricht auch, dass, ähnlich wie bei Windows Media mit 192 kbps, knapp in der Hälfte aller Fälle 90 Punkte und mehr vergeben wurde. Mit der zweitniedrigsten Standardabweichung scheint auch Ogg Vorbis homogene Qualität zu erreichen. Zwar gibt es auch hier Ausreißer nach unten, die den Mittelwert gegenüber dem Median deutlich abschwächen, jedoch weniger stark als bei Windows Media 192 kbps.

**Dolby Digital mit 448 kbps** war anscheinend mit 25% aller Fälle über 90 Punkten, 75% aller Werte über 50 Punkten und Median und Mittelwert über 70 Punkten für die Testpersonen ebenfalls ein gutes Verfahren, das aber möglicherweise bei einigen Signaltypen Probleme aufweist. Das wird in der weiteren Auswertung zu beobachten sein.



Für die **versteckte Referenz** ist zu erkennen, dass keine Wertungen unterhalb von 80 Punkten zugelassen wurden und sich somit alle Werte im Excellent Bereich befinden. Mehr als 50% wurde mit 100 Punkten bewertet.

Der **Anchor** befindet sich im Bad Bereich, Wertungen, die darüber hinaus gehen würden, sind von der Auswertung ausgenommen worden. Etwa 95% aller Bewertungen sind kleiner als 10 Punkte ausgefallen.

## 7.2 Darstellung der Mittelwerte und Konfidenzintervalle nach Kodierverfahren und Hörbeispielen aufgeteilt

Wie in der Analyse des Boxplots gesehen, werden die Verfahren teils deutlich unterschiedlich, teils aber auch recht ähnlich bewertet, wie beispielsweise bei Windows Media 192 kbps und Dolby Digital 448 kbps. Ob sich vermeintliche Ähnlichkeiten oder aber Unterschiede statistisch auch als signifikant herausstellen, soll nun mittels eines Tests der Mittelwerte auf Gleichheit, wie in der ITU Empfehlung beschrieben, untersucht werden.

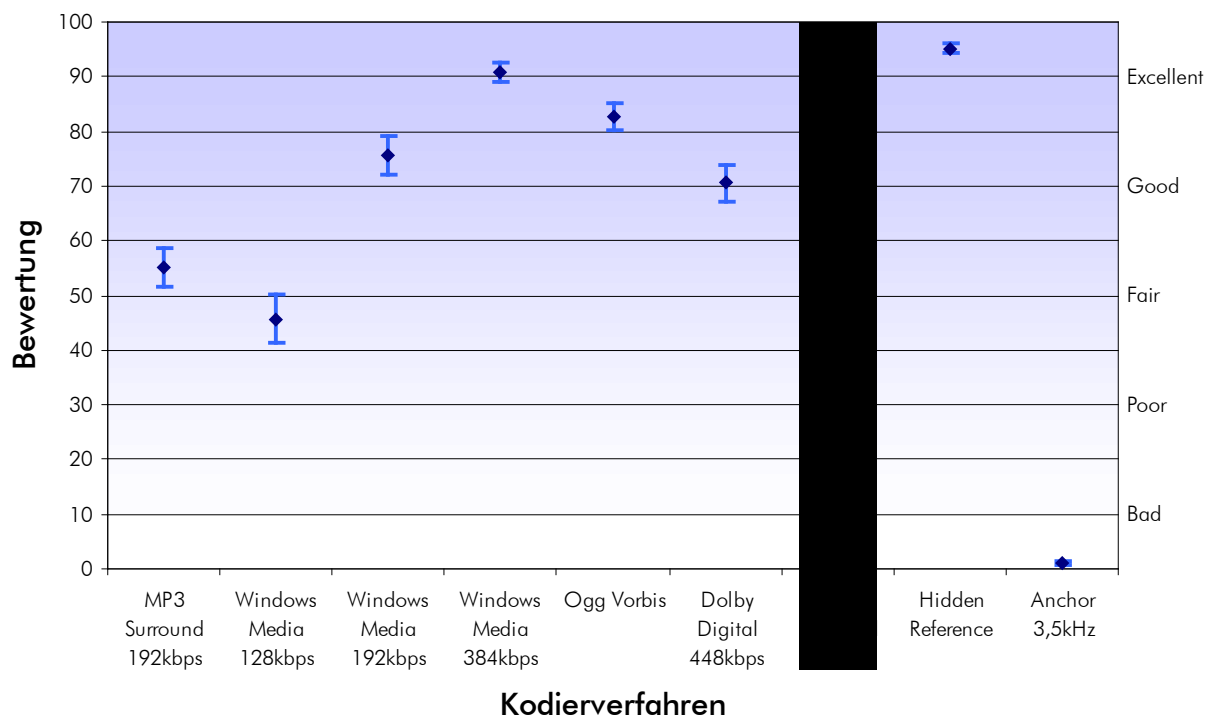


Für die Darstellung der errechneten Werte wurden folgende Symbole gewählt:

- ◆ kennzeichnet den errechneten arithmetischen Mittelwert

I bezeichnet den Bereich des Konfidenzintervalls in der zu 95% der tatsächliche Wert liegt

### 7.2.1 Alle Kodierverfahren über alle Hörbeispiele



Wie aus der Darstellung der Mittelwerte und Konfidenzintervalle ersichtlich wird, scheint Windows Media 384 kbps für die Testhörer am besten geklungen zu haben, danach folgt Ogg Vorbis, Windows Media 192 kbps und Dolby Digital 448 kbps. Die Konfidenzintervalle der ersten drei überlappen nicht, deshalb kann man davon ausgehen, dass es signifikante Unterschiede zwischen den Kodierverfahren gibt. Anders bei den letzten beiden, wo eine knappe Überlappung auszumachen ist, was bedeuten würde, dass man keine signifikanten Unterschiede feststellen konnte. Diese Erkenntnisse möchte ich anhand des Wilcoxon-Vorzeichen-Rang-Tests absichern und die Nullhypothese der Gleichheit der Werte überprüfen. Das Signifikanzniveau habe ich dabei auf 95% festgelegt.

Nacheinander stelle ich nun die folgenden Kodierverfahren gegenüber:

Windows Media 384 kbps - Ogg Vorbis

Windows Media 384 kbps - Windows Media 192 kbps

Windows Media 192 kbps - Ogg Vorbis

Windows Media 192 kbps - Dolby Digital 448 kbps

Der Test ergibt für die geprüften Fälle, dass die Annahme der Gleichheit der durchschnittlichen Bewertungen widerlegt werden kann. Der Wilcoxon-Vorzeichen-Rang-Test gibt in den ersten drei Fällen der Grafik recht, im letzten Fall widerlegt er die grafische Auswertung. Somit wäre festzustellen, dass unter den vier am höchsten bewerteten Kodierverfahren, deren Mittelwerte mindestens im Good Bereich liegen, Windows Media 384 kbps insgesamt am besten abgeschnitten hat, vor Ogg Vorbis und Windows Media 192 kbps. Dolby Digital mit 448 kbps und damit der höchsten Datenrate, wurde von diesen vier Verfahren insgesamt offensichtlich am schlechtesten bewertet.

Zusätzlich möchte ich prüfen, ob Windows Media 384 kbps tatsächlich signifikante Unterschiede zum Original, also zur Bewertung der Hidden Reference aufweist:

Windows Media 384 kbps - Hidden Reference

Die Nullhypothese wird auch hier nicht bestätigt, der Test weist signifikante Unterschiede zum gewählten Niveau auf. Der Grafik kann weiterhin Recht gegeben werden.

Ähnlich will ich nun bei den nächsten [REDACTED] Kodierverfahren vorgehen, die erkennbar von den oberen vier abgesetzt sind. Um sie jedoch besser einordnen zu können, führe ich auch hier untereinander Wilcoxon-Vorzeichen-Rang-Tests durch. Wie die Analyse mithilfe der Konfidenzintervalle zeigt, weisen MP3 Surround 192 kbps und Windows Media 128 kbps offenbar signifikante Unterschiede auf [REDACTED]

[REDACTED]:

MP3 Surround 192 kbps - Windows Media 128 kbps

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] Die Nullhypothese wird [REDACTED]

[REDACTED] verworfen [REDACTED]

[REDACTED]. Der Wilcoxon-Vorzeichen-Rang-Test zeigt somit an dieser Stelle signifikante Unterschiede im Vergleich zwischen Windows Media 128 kbps zu MP3 Surround [REDACTED]

[REDACTED] auf [REDACTED]

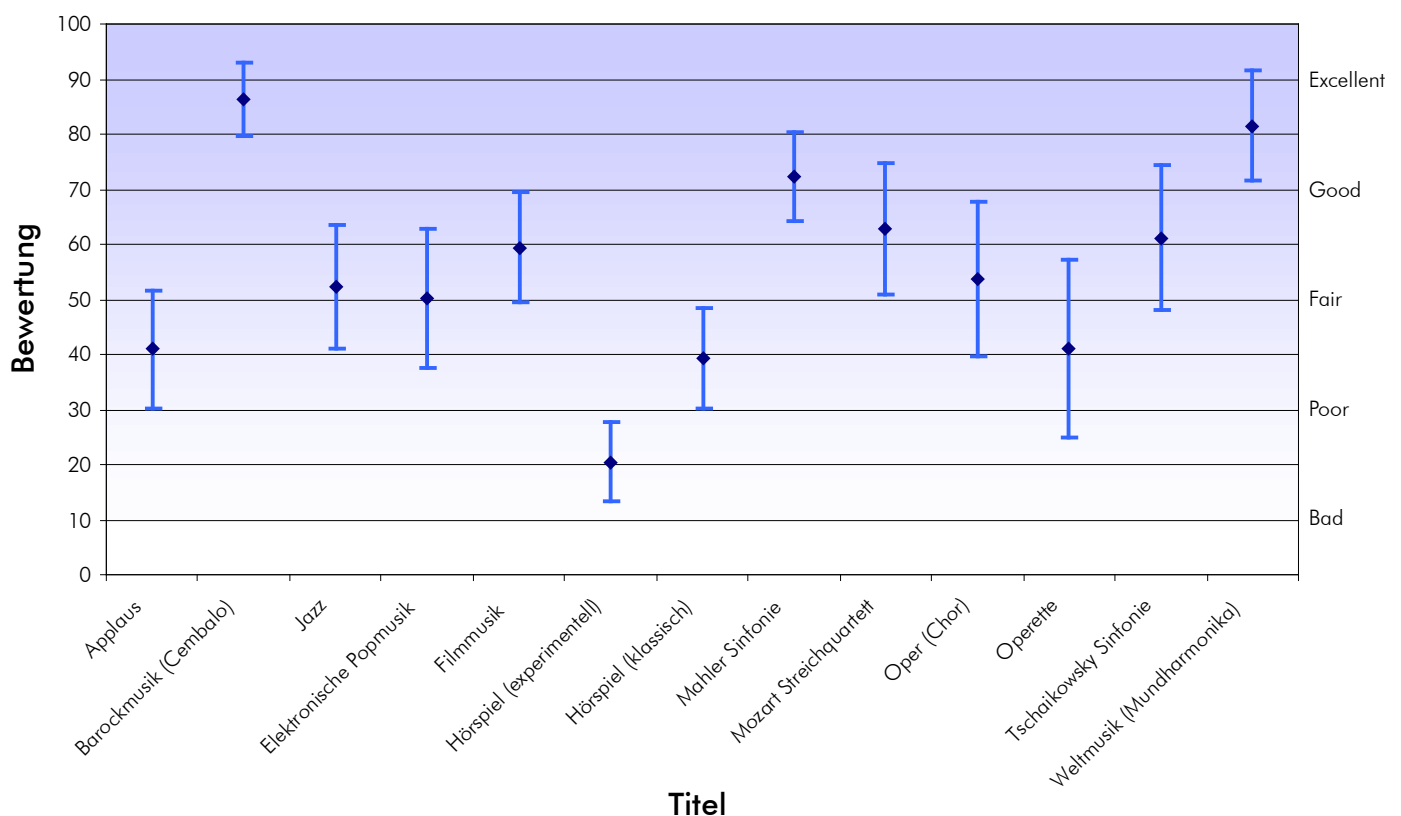
[REDACTED]. Die Feststellung bei der Betrachtung der Grafik, dass Windows Media 128 kbps im Mittel am niedrigsten bewertet wurde [REDACTED]

[REDACTED] kann erhalten bleiben.

## 7.2.2 Mittelwerte und Konfidenzintervalle aller Hörbeispiele separat für jedes Kodierverfahren

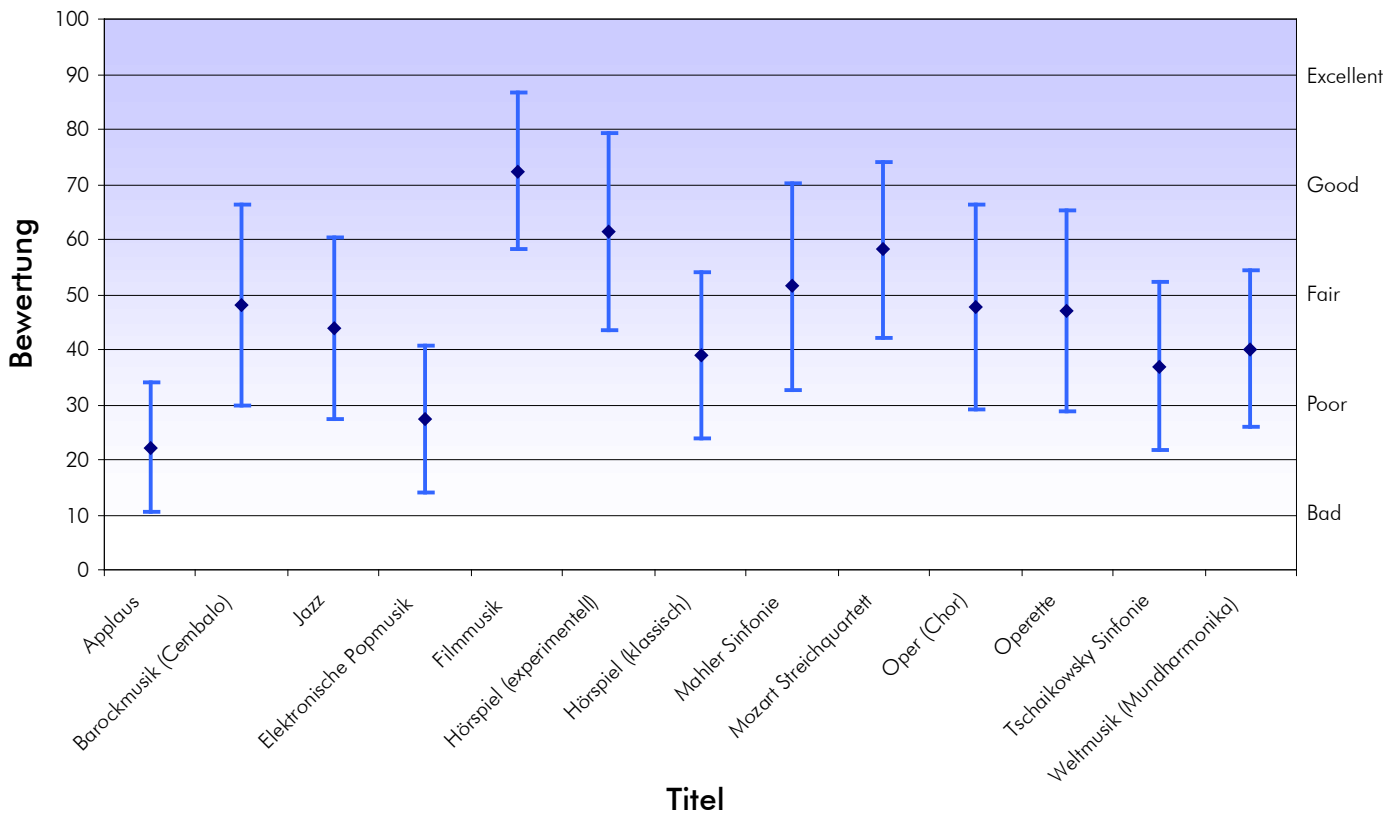
Als nächstes werden die Mittelwerte und Konfidenzintervalle der einzelnen Hörbeispiele für jedes Kodierverfahren näher betrachtet. Da es sich dabei allerdings nicht mehr um paarweise durchgeführte Tests handelt, muss auf die Überprüfung der Ergebnisse mit dem Wilcoxon-Vorzeichen-Rang-Test verzichtet werden. Dennoch können Tendenzen abgelesen werden.

### Mittelwerte aller Hörbeispiele für Kodierverfahren MP3 Surround 192 kbps



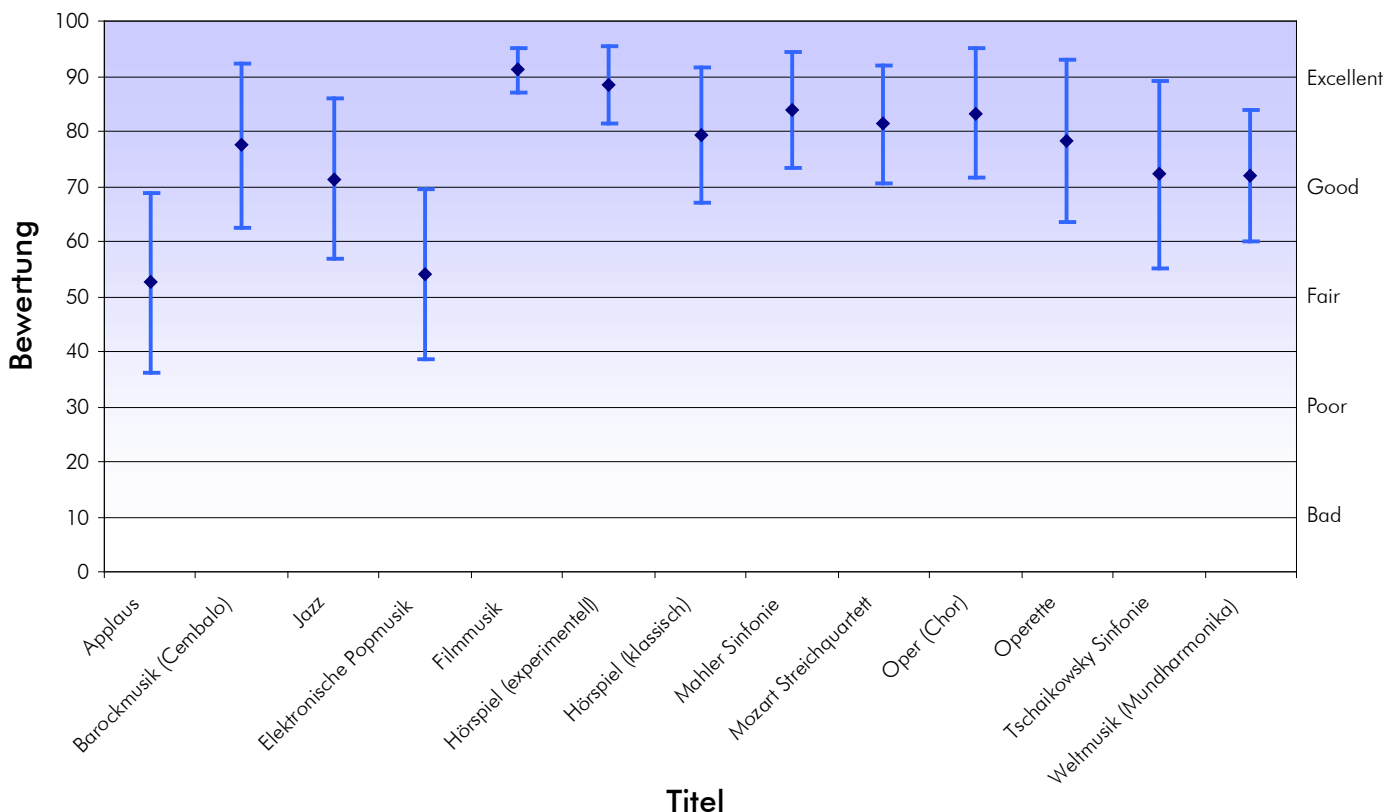
Wie bereits bei der Betrachtung des Boxplots vermutet wurde, weist MP3 Surround sehr große Unterschiede je nach Art des gewählten Beispiels auf. Die errechneten Mittelwerte erreichen nicht einmal 40 Punkte bei den beiden Hörspielausschnitten. Das Barockstück dagegen wurde im Mittel am besten bewertet, obwohl sich in früheren Tests Musik mit Cembalo häufig als kritisch herausgestellt hatte. Gleiches gilt auch für das Beispiel mit der Mundharmonika, beide Mittelwerte liegen im Excellent Bereich. Dagegen scheint dieses Verfahren mit 192 kbps bei Applaus weiterhin nicht sehr überzeugend zu sein. Alle weiteren Ausschnitte liegen im Mittel im Fair oder Good Bereich wobei in den wenigsten Fällen von signifikanten Unterschieden ausgegangen werden kann, da die Konfidenzintervalle überwiegend überlappen.

### Mittelwerte aller Hörbeispiele für Kodierverfahren Windows Media 128 kbps



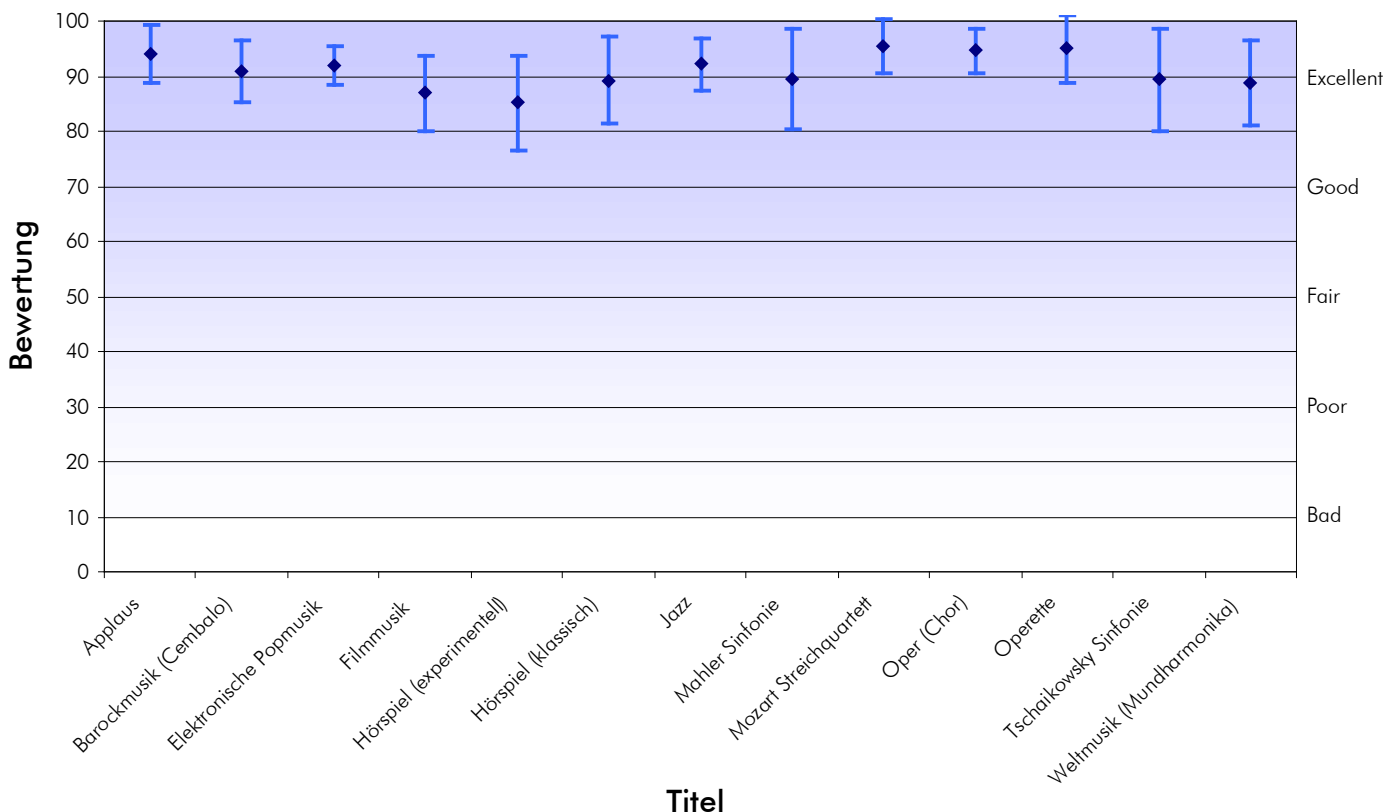
Bei Windows Media 128 kbps kann man feststellen, dass die Mittelwerte für elf der 13 Fälle unter 60 Punkte bleiben, wobei Applaus am schlechtesten abschneidet. Lediglich das Beispiel für Filmmusik sowie das experimentelle Hörspiel, dass bei MP3 Surround noch als Problemfall angesehen werden konnte, scheinen im Mittel gute Ergebnisse zu liefern. Eine große Streuung, die bei der Analyse des Boxplots bereits angemerkt wurde, scheint sich durch die im Vergleich zu allen anderen Formaten relativ ausgedehnten Konfidenzintervalle zu bestätigen. Obwohl insgesamt bei Windows Media 128 kbps nicht einmal die Hälfte aller Bewertungen über 40 Punkte hinausreichte (siehe Boxplotgrafik S. 38), liegen die Mittelwerte in acht Fällen über dieser Marke, was als weiteres Indiz häufiger Ausreißer nach oben gewertet werden darf. Überwiegend überlappende Konfidenzintervalle geben Grund zur Annahme, dass in den meisten Fällen qualitativ ähnliche Ergebnisse erzielt werden.

### Mittelwerte aller Hörbeispiele für Kodierverfahren Windows Media 192 kbps



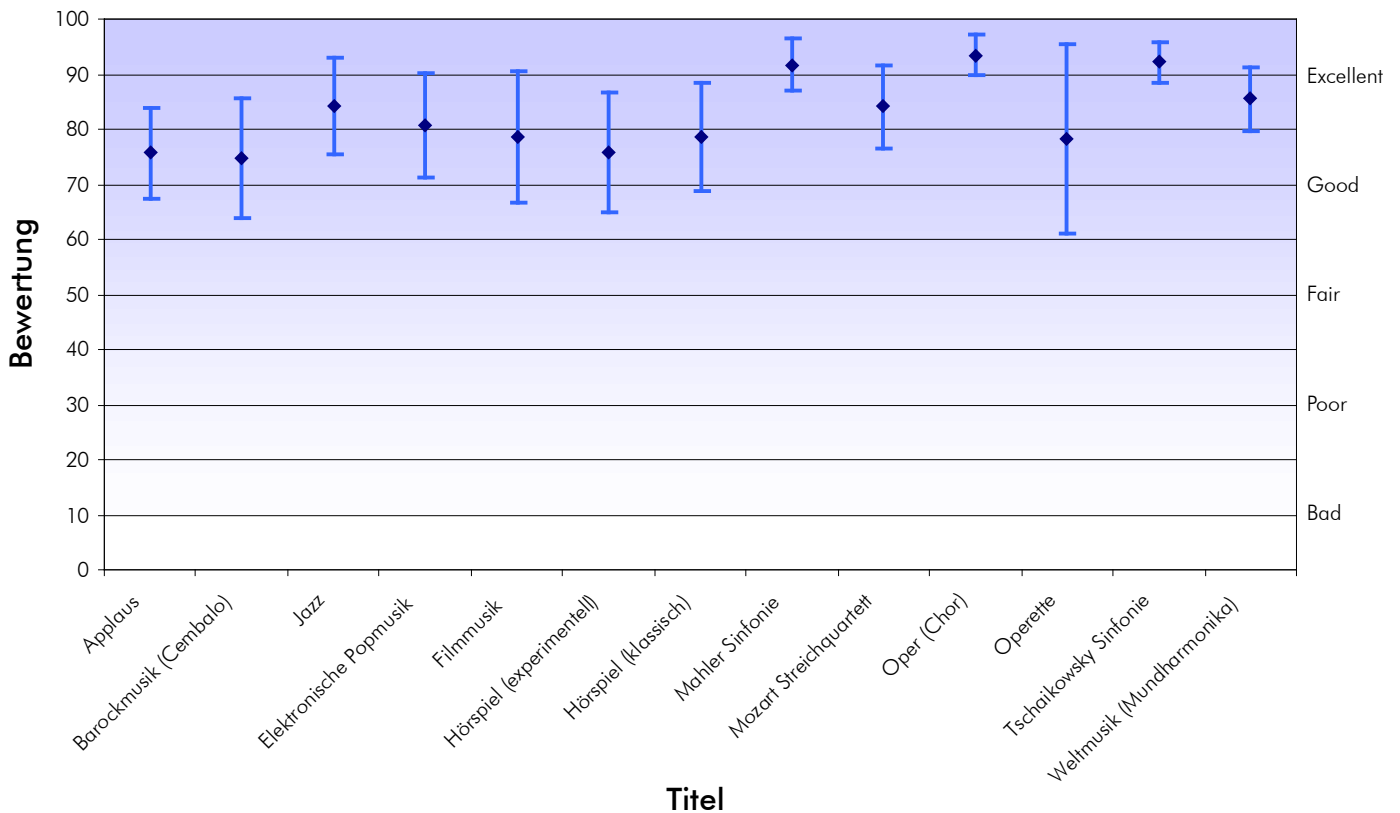
Deutlich besser wurde Windows Media 192 kbps empfunden, denn die Mittelwerte von elf der 13 Hörbeispiele liegen bei über 70 Punkten. Das deutet auch auf eine größere Homogenität in der Qualität der Kodierung hin. Interessant zu beobachten ist, dass die Ausschnitte für Applaus und Elektronische Popmusik, wie schon beim gleichen Verfahren mit 128 kbps, die niedrigsten Mittelwerte aufweisen. Vermutlich findet sich hier der Grund für die Abweichung des Mittelwerts zum Median (siehe Boxplotanalyse), denn in diesen beiden Fällen setzen sich die Mittelwertsmarkierungen am deutlichsten von den restlichen Ergebnissen aus den Bewertungen ab. Ebenfalls findet sich die Tendenz wieder, dass die Filmmusik im Mittel am besten eingestuft wurde, daraus kann man Parallelitäten bei der Anwendung des Verfahrens mit verschiedenen Datenraten ableiten, zumindest bis 192 kbps.

### Mittelwerte aller Hörbeispiele für Kodierverfahren Windows Media 384 kbps



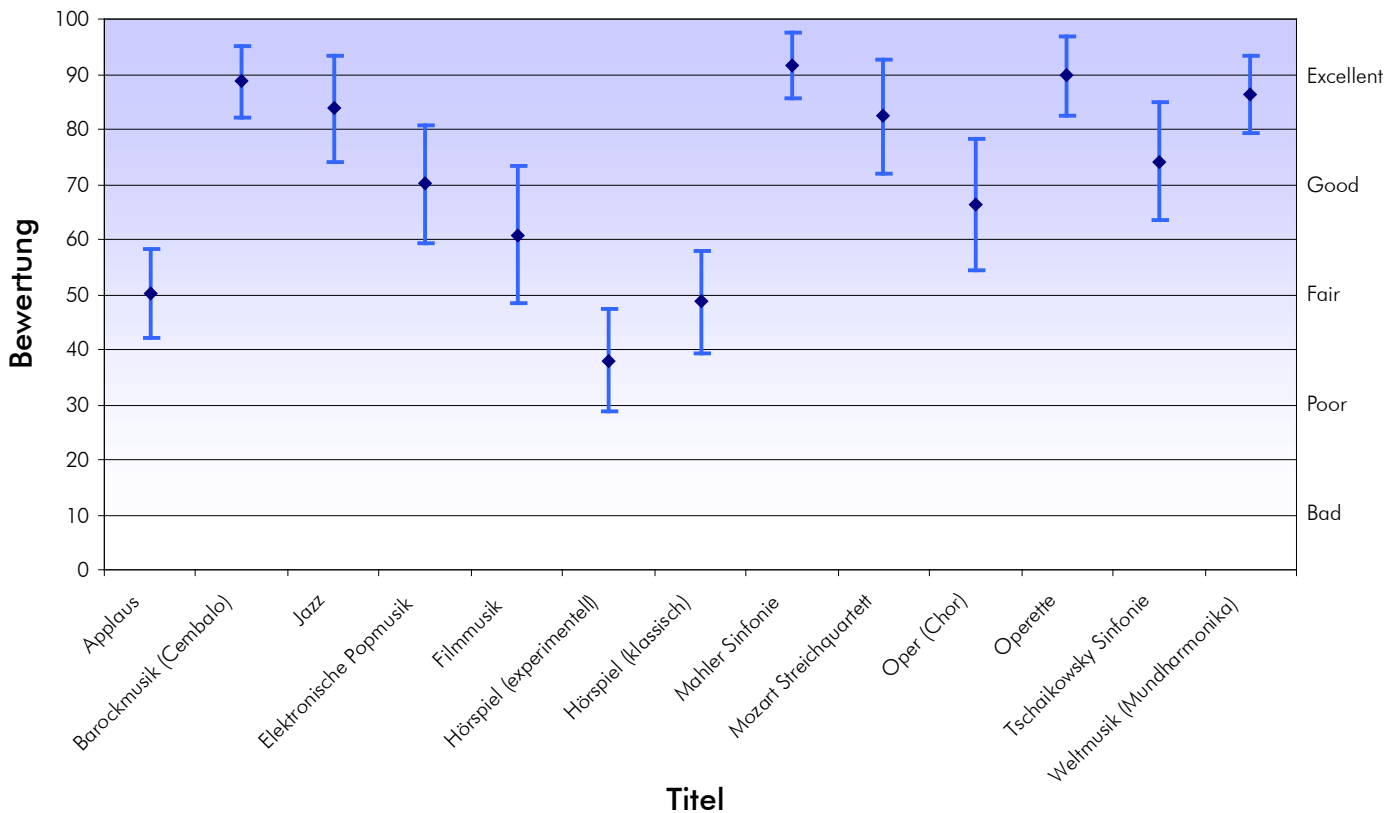
Windows Media 384 kbps liegt bei allen Hörbeispielen im Mittel im Excellent Bereich. Zudem kann man annehmen, dass zwischen den einzelnen Audioausschnitten keine signifikanten Unterschiede auszumachen sind, da alle Konfidenzintervalle untereinander überlappen. Selbst die Auffälligkeiten, die noch bei Bitraten von 128 kbps und 192 kbps aufgetreten sind finden sich bei der deutlich höheren Datenrate nicht wieder. Insgesamt scheint die Qualität als sehr gut und konstant beurteilt worden zu sein. Wo die im Boxplot ausgemachten Ausreißer nach unten angesiedelt sind, lässt sich nicht erkennen. Sie scheinen in verschiedenen Beispielen aufzutreten, vermutlich in den beiden Hörspielausschnitten und den beiden Sinfonien, da die Vertrauensintervalle am meisten Punkte umfassen.

### Mittelwerte aller Hörbeispiele für Kodierverfahren Ogg Vorbis



In der Boxplotgrafik war bereits zu erkennen, dass auch Ogg Vorbis konstant gute und sehr gute Ergebnisse zu liefern scheint. In zehn von 13 Fällen schwankt der errechnete Mittelwert in dem Bereich, zwischen 74,72 und 85,5 Punkten. Lediglich die Sinfoniausschnitte und die Oper setzen sich im Mittel nach oben hin etwas ab, mutmaßlich war auch die Streuung in diesen Fällen geringer, dieser Eindruck wird zumindest durch den Umfang der Konfidenzintervalle erweckt. Im Gegensatz dazu scheinen bei der Operette ganz unterschiedliche Meinungen vorhanden gewesen zu sein, möglicherweise gibt es auch einige Ausreißer nach unten, denn der Bereich, in dem zu 95% der tatsächliche Mittelwert liegt, umfasst weitaus am meisten Punkte.

### Mittelwerte aller Hörbeispiele für Kodierverfahren Dolby Digital 448 kbps



Dolby Digital, das bisher einzige schon eingesetzte Verfahren, liefert erkennbar differierendere Ergebnisse als die drei zuletzt analysierten. Zwar scheint Musik im Schnitt gut oder sehr gut kodiert werden zu können (in sechs Fällen ergibt der Mittelwert über 80 Punkte), dennoch scheinen die Hörspielstellen und auch der Applaus im Vergleich dazu das System an die Grenzen gebracht zu haben, denn mit 95%iger Sicherheit liegt der tatsächliche Mittelwert wohl unterhalb des Good Bereichs. Obwohl bei diesem Kodierverfahren die höchste Datenrate angewendet wurde, macht es den Anschein, als könne nicht immer eine hohe Qualität gewährleistet werden. Für die bisherige Anwendung bei der Übertragung von klassischer Musik über DVB-S liefert es wohl sehr ordentliche Ergebnisse, die Qualitätsminderung bei Applaus scheint man dabei billigend in Kauf nehmen zu können. Wollte man allerdings, wie es für Bayern 2 denkbar wäre, in Zukunft auch Hörspiel im Mehrkanalformat übertragen, so bedarf es kritischer Beobachtung, in wie weit dann Qualitätsminderungen auszumachen sein könnten.



[Redacted]

[Redacted]

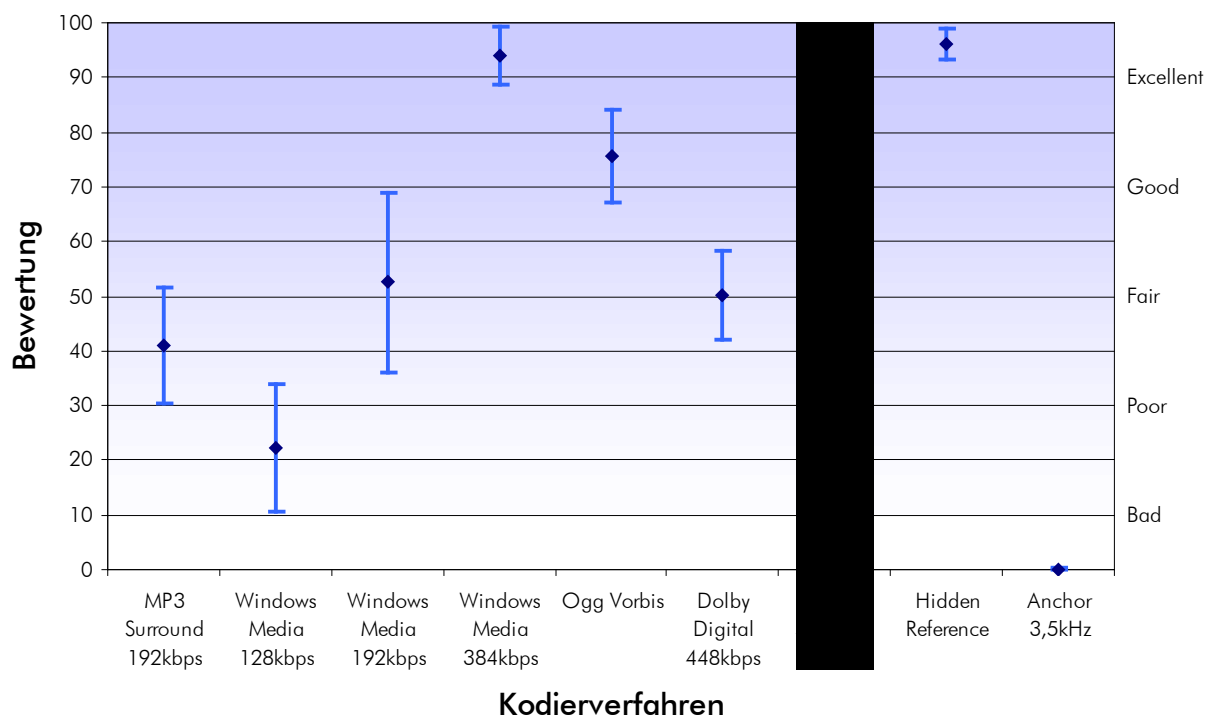
[Redacted]

### 7.2.3 Ergänzende Darstellung der Mittelwerte und Konfidenzintervalle aller Kodierverfahren separat für jedes Hörbeispiel

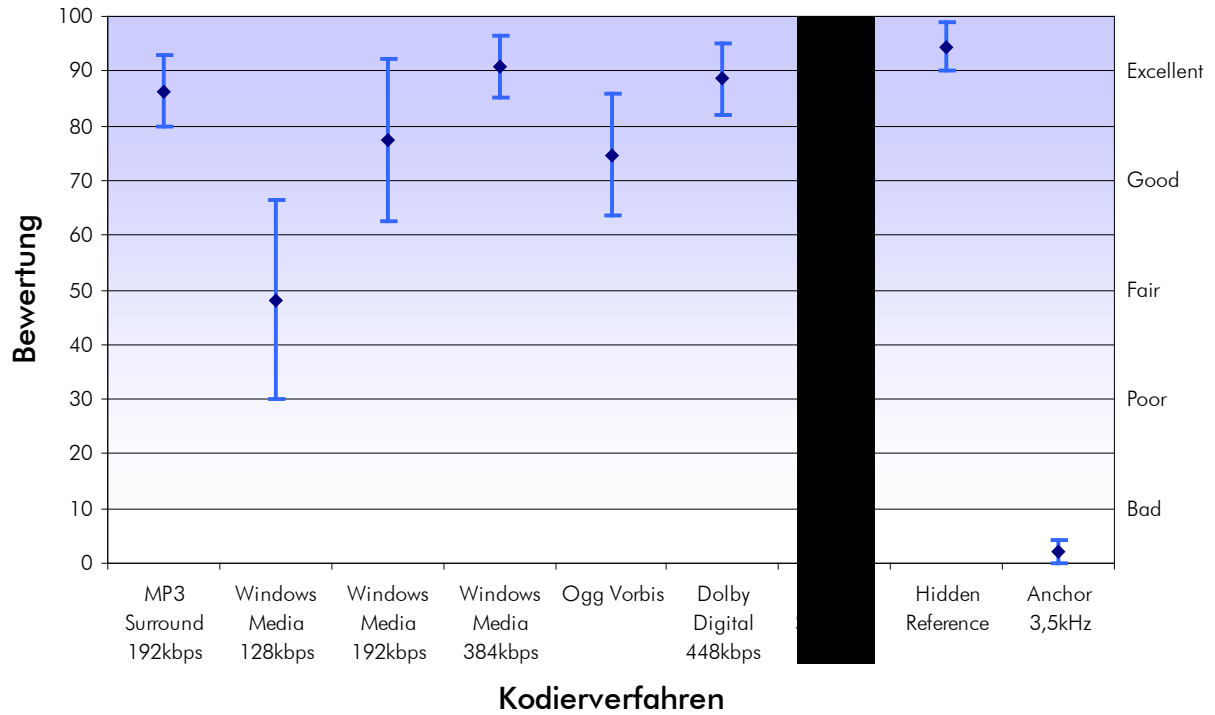
Die folgenden Grafiken stellen die einzelnen Kodierverfahren für das jeweilige Hörbeispiel nochmals gegenüber und geben anhand der eingezeichneten Vertrauensintervalle nochmals Aufschluss darüber, wo signifikante Unterschiede zwischen den Verfahren in den einzelnen Audioausschnitten zu finden sind.

Besonders interessant ist diese Darstellung für Signalarten, bei denen die Bewertung unterdurchschnittlich für ein bestimmtes Kodierverfahren war, beispielsweise für Applaus bei Windows Media mit 128 kbps und 192 kbps sowie bei Dolby Digital.

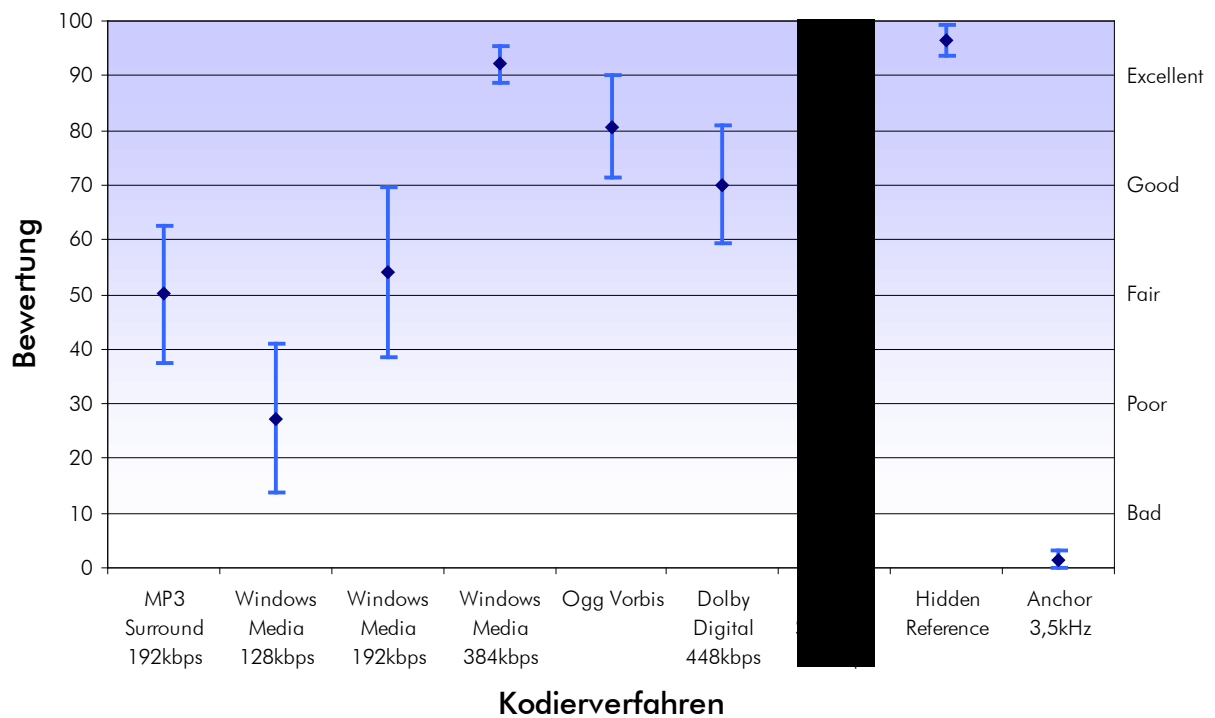
#### Alle Kodierverfahren für Hörbeispiel Applaus



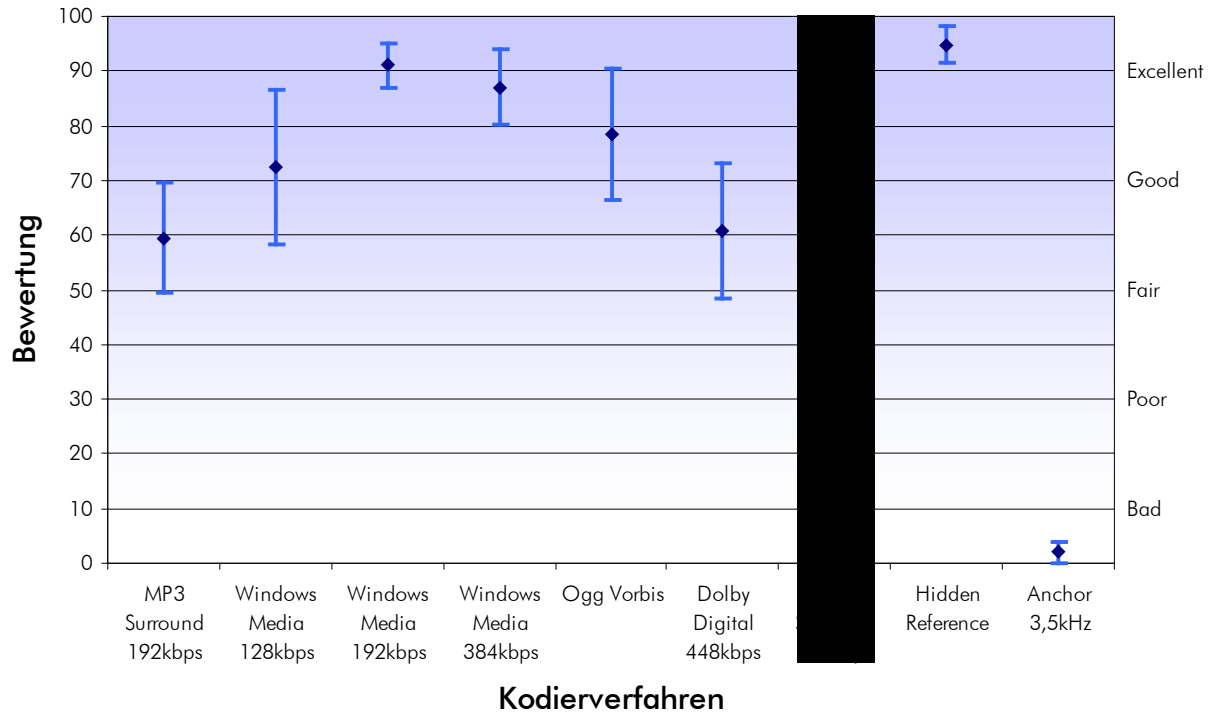
### Alle Kodierverfahren für Hörbeispiel J. C. Bach (Barockmusik mit Cembalo)



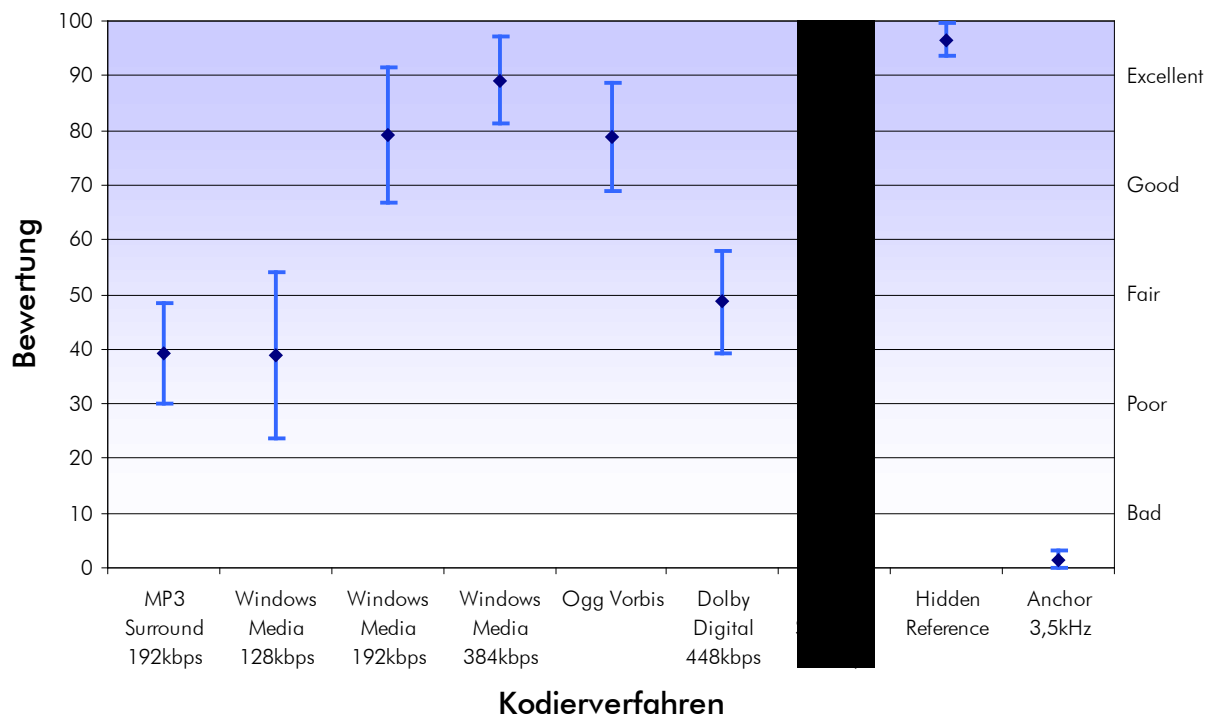
### Alle Kodierverfahren für Hörbeispiel Crystal Method (Elektronische Popmusik)



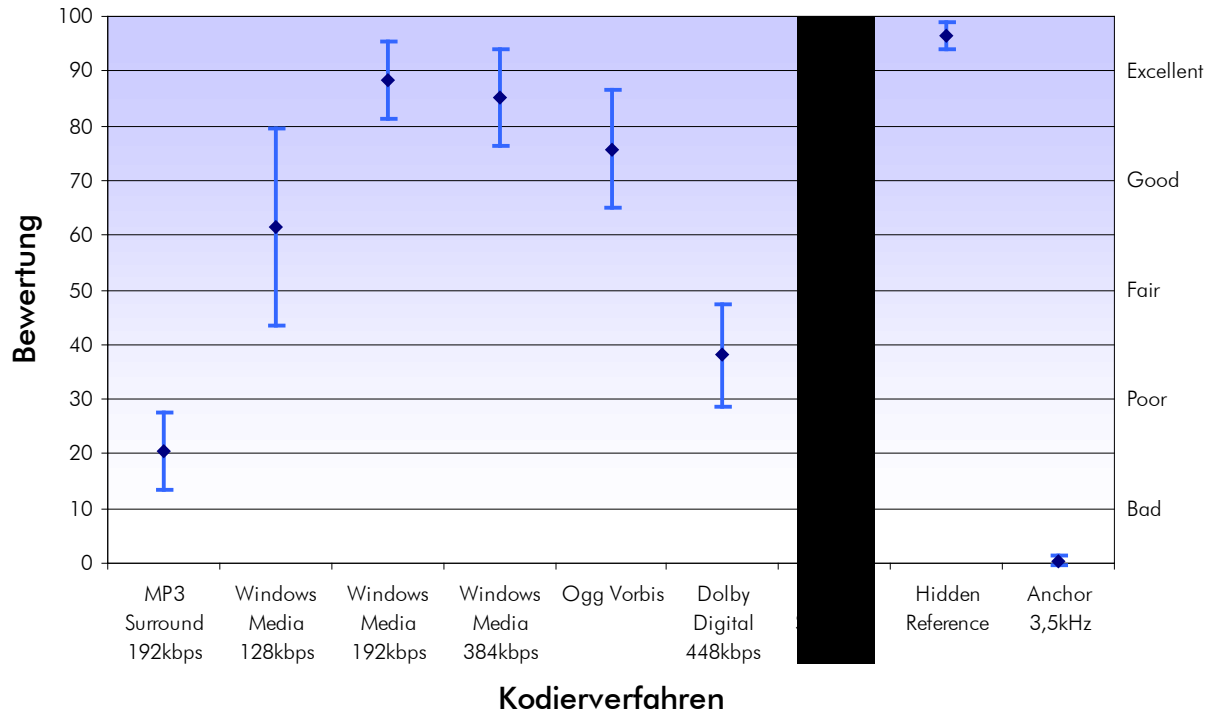
### Alle Kodierverfahren für Hörbeispiel Love Story (Orchestrale Filmmusik)



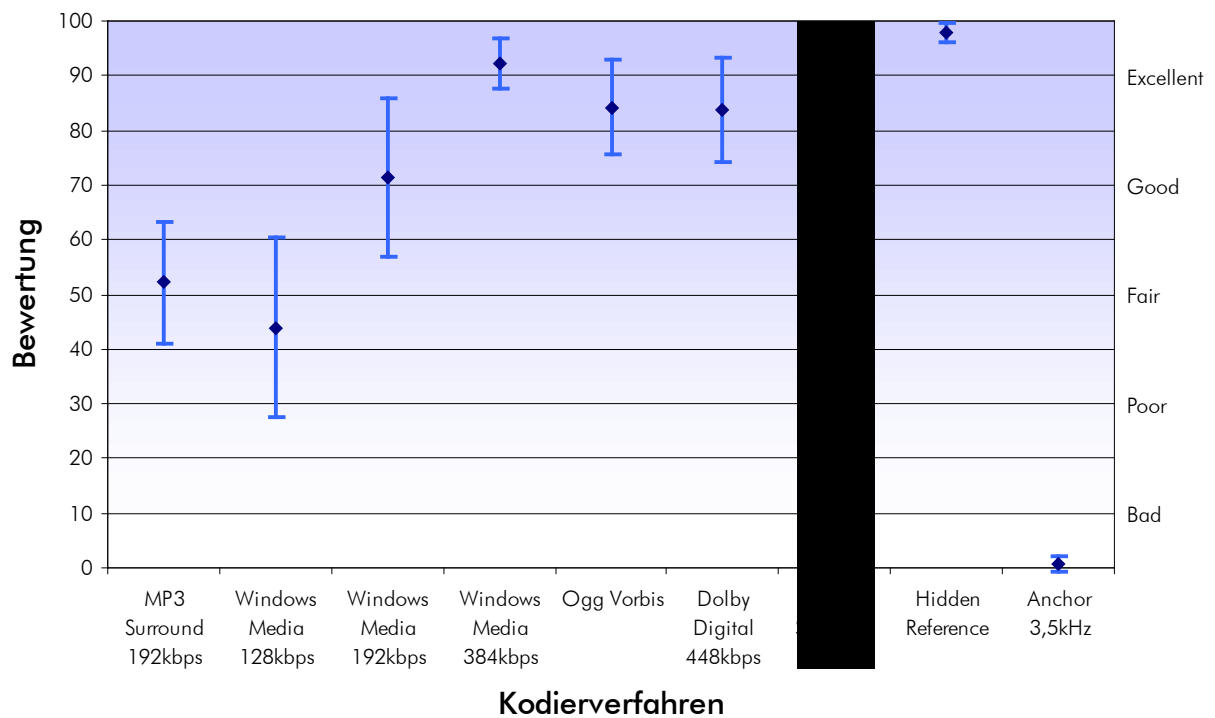
### Alle Kodierverfahren für Hörbeispiel Jules Verne (Hörspiel)



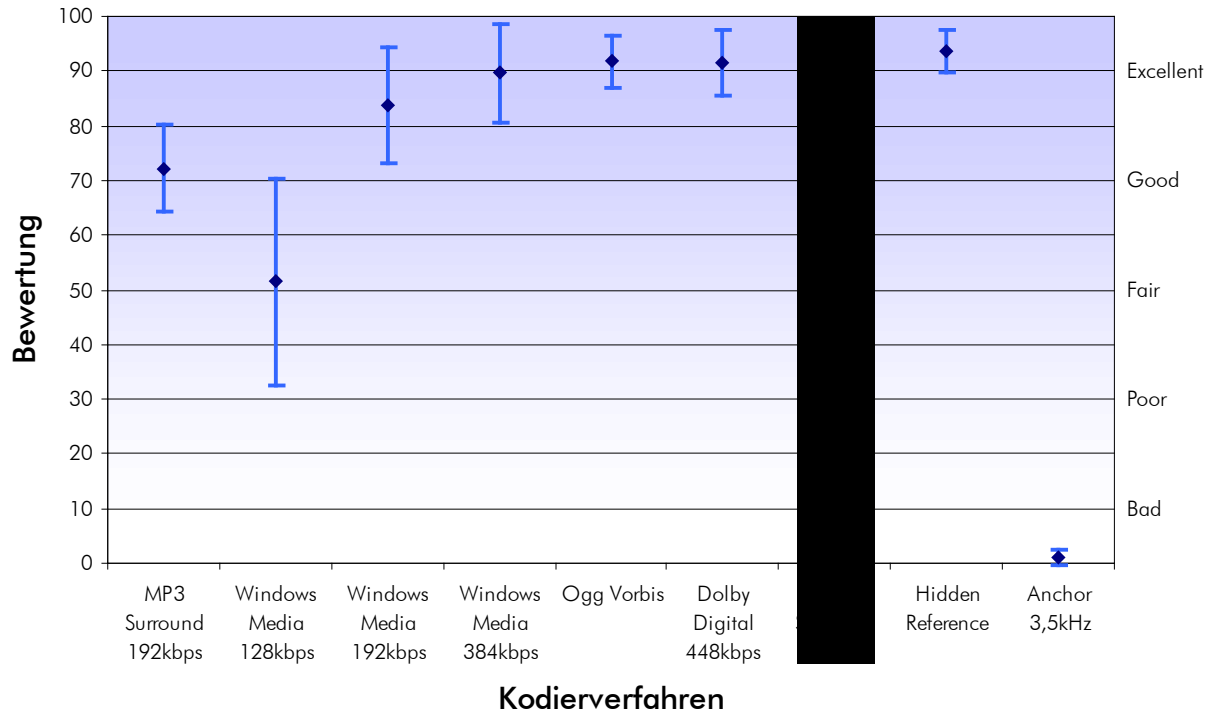
### Alle Kodierverfahren bei Hörbeispiel Wunschzeit (experimentelles Hörspiel)



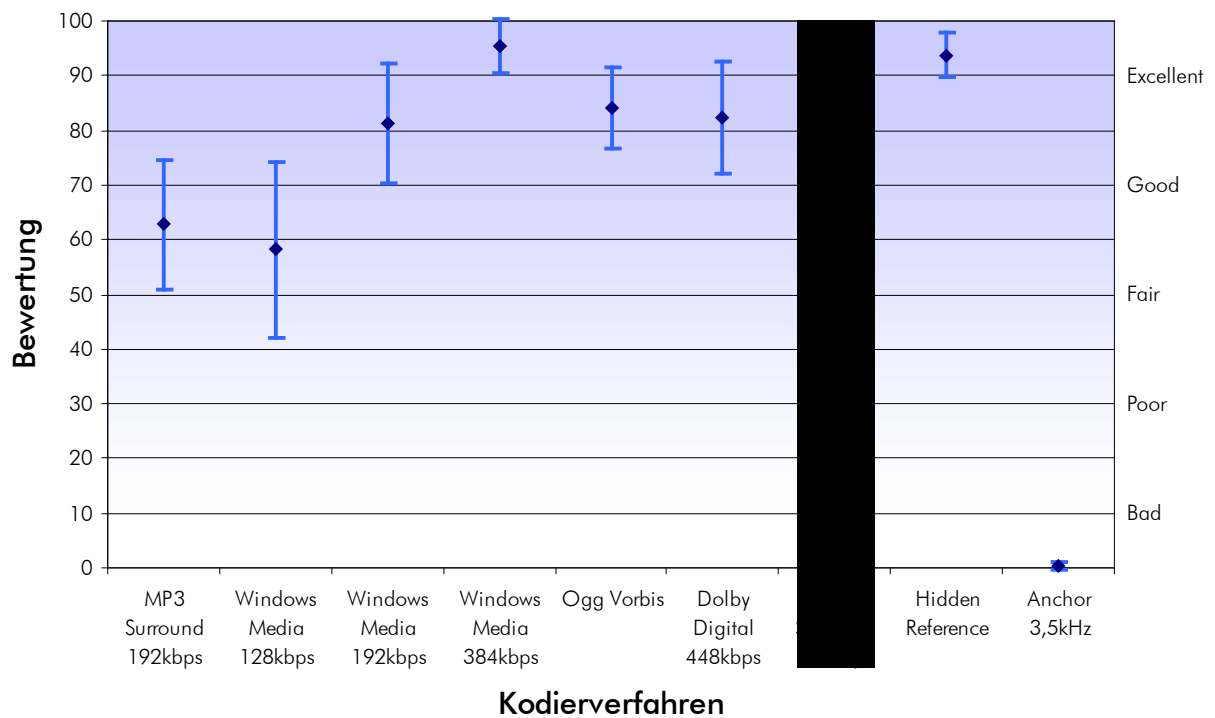
### Alle Kodierverfahren für Hörbeispiel Leuchter (Jazz)



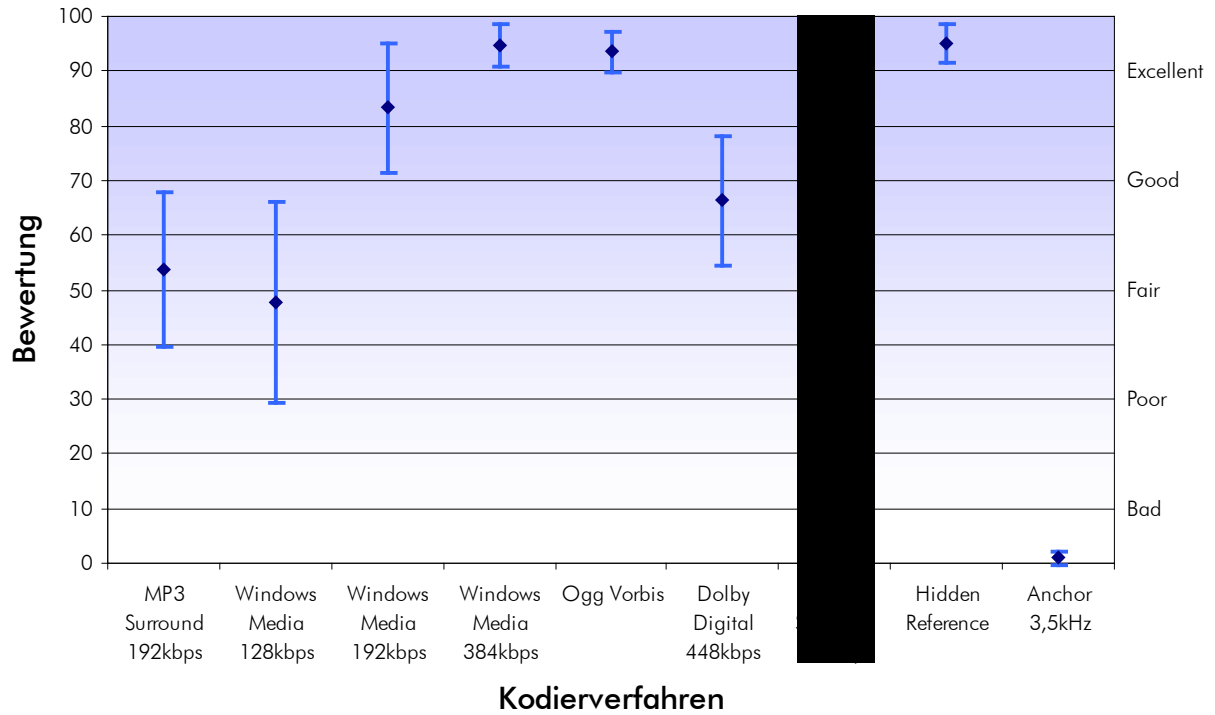
### Alle Kodierverfahren für Hörbeispiel Mahler (Klassische Musik mit großem Orchester)



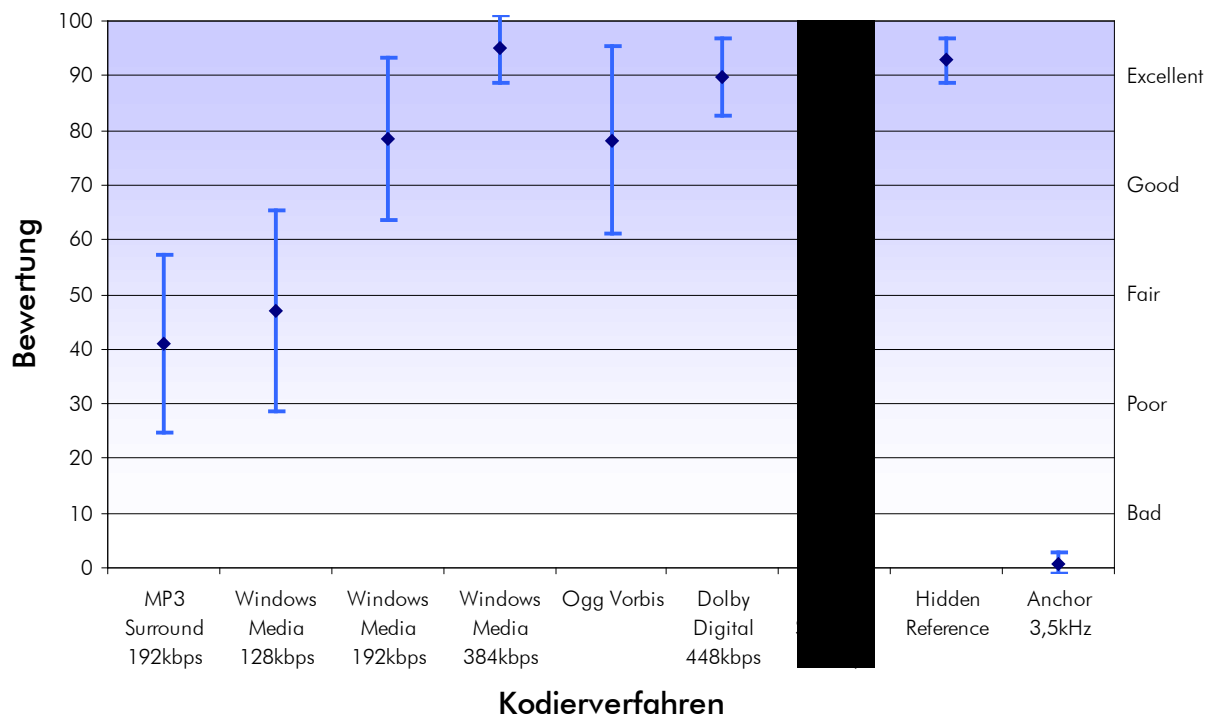
### Alle Kodierverfahren bei Hörbeispiel Mozart (Streichquartett)



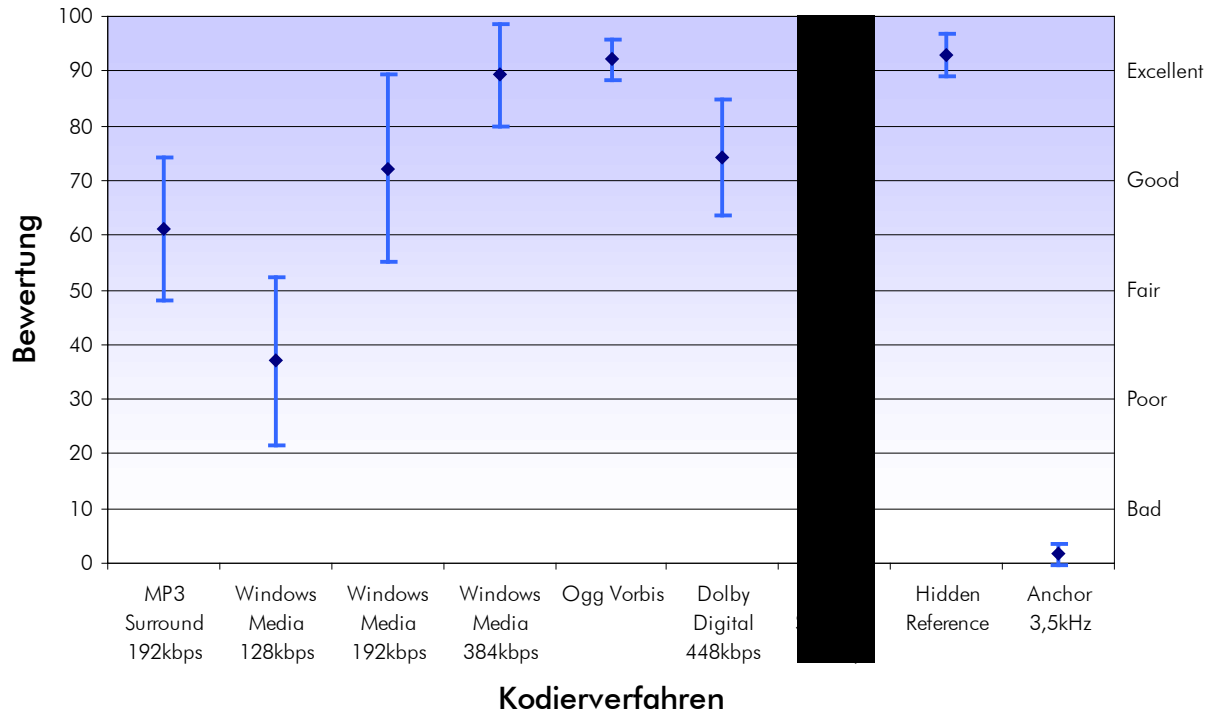
### Alle Kodierverfahren bei Hörbeispiel Nabucco (Chor in Oper)



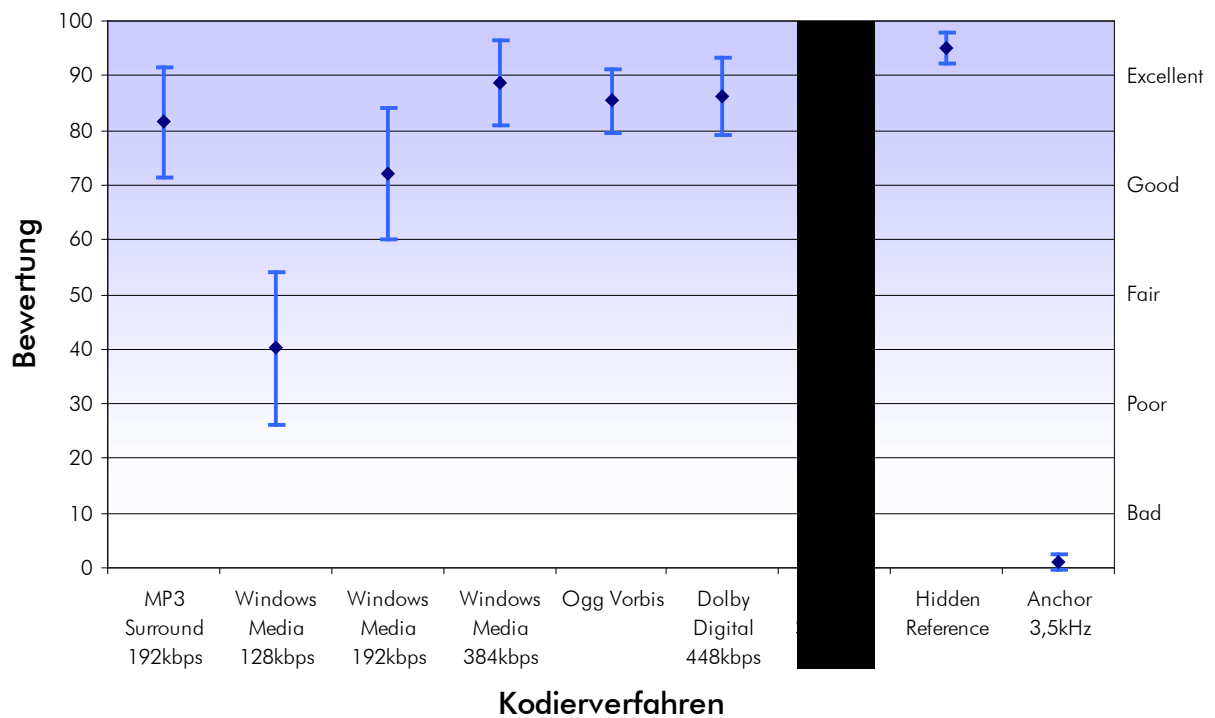
### Alle Kodierverfahren bei Hörbeispiel Wiener Blut (Operette)



### Alle Kodierverfahren für Hörbeispiel Tchaikowsky (Klassische Musik mit großem Orchester)



### Alle Kodierverfahren bei Hörbeispiel Maná (Weltmusik mit Mundharmonika)





Bei Stücken, die in der klassischen Musik angesiedelt sind, haben die vier durchschnittlich am besten bewerteten Verfahren (Windows Media 192 kbps und 384 kbps, Ogg Vorbis, Dolby Digital 448kbps) insgesamt recht gute Ergebnisse erzielt; sie lagen stets mindestens im Good Bereich, wohingegen die restlichen Verfahren zum Teil durchwachsene Resultate für den Mittelwert lieferten.

Wie sich zeigte, haben die vermeintlich kritischen Signaltypen Cembalo und Mundharmonika keine unterdurchschnittlichen Ergebnisse bei den jeweiligen Kodierverfahren geliefert.

Anders stellt sich die Sache bei Applaus dar, wo die Reihenfolge der durchschnittlichen Bewertung der Reihenfolge der Mittelwerte und Mediane über alle Materialien in 7.1 entspricht. Applaus scheint deshalb ein Indikator für die Qualität eines Kodierverfahrens zu sein. Zu beachten ist allerdings, dass die Mittelwerte für Applaus unter den Gesamtmittelwerten liegen (einzige Ausnahme bildet Windows Media 384 kbps) und somit also nur die Tendenz als aussagekräftig angesehen werden kann.

Auffällig waren zum Teil die Hörspielausschnitte. Hier konnten sich die Verfahren, deren Mediane und Gesamtmittelwerte die höchsten Werte erzielten, zum Teil erheblich von den übrigen absetzen. Es ist zu vermuten, dass je nach Machart, einige Kodierverfahren Probleme mit solchen Produktionen haben.

## 8. Fazit

Sollte der Bayerische Rundfunk zukünftig sein Mehrkanalangebot auch über das Internet verbreiten wollen, würde ich empfehlen, hierfür Ogg Vorbis bzw. Windows Media mit 192 kbps oder höher in Betracht zu ziehen.

MP3 Surround drängt sich zwar als Mehrkanalpendant zu dem sehr beliebten MP3 Format für Mono- und Stereoanwendungen auf, liefert allerdings selbst bei 192 kbps keine sehr stabile Qualität, wie die Analyse der Tests zeigte.

Windows Media (128 kbps) scheint bei der gleichen Datenrate, die auch für den Stereostream der meisten Programme derzeit verwendet wird, zu schlechte Ergebnisse zu liefern. Deshalb müsste bei der Verwendung dieses Verfahrens eine Datenrate von mindestens 192 kbps angestrebt werden. Die von mir getestete Datenrate von 384 kbps liefert scheinbar sogar nicht nur bessere, sondern auch konstantere Qualität als das für DVB-S verwendete Dolby Digital bei 448 kbps. Da die erforderliche Bandbreite heutzutage in vielen Haushalten bereits zur Verfügung steht und somit kein Problem darstellen sollte, wäre ein entsprechendes Mehrkanalangebot des Bayerischen Rundfunks auf Basis des Internetstreamings wünschenswert.

Als Alternative könnte auch das Ogg Vorbis Verfahren infrage kommen. Bisher spielt es im Angebot des BR keine Rolle, es liefert jedoch ebenfalls konstant gute Ergebnisse beim Encoding und wird auch von vielen Playern unterstützt (z. B. Winamp, VLC oder Windows Media Player). Die Verwendung variabler Bitraten scheint sehr gut zu funktionieren, selbst bei niedrigen Einstellungen in der Qualitätsabstufung. Der Bandbreitenbedarf könnte damit sogar geringer sein als bei Windows Media mit 192 kbps.

Ich denke, dass eine Ausweitung des Mehrkanalangebots in Zukunft ein Thema im Hörfunk sein wird, gerade im Klassiksegment, wo auch immer mehr Produktionen auf den Markt kommen. Sollte der Bayerische Rundfunk in diesem Bereich aktiv werden, ist auch das Internetstreaming fast schon Pflicht. Deshalb hoffe ich, dass ich meinen Beitrag dazu leisten konnte und die Entscheidung zu diesem Schritt leichter fällt.

## Literaturverzeichnis

**Bayerischer Rundfunk (Hrsg.):** Lexikon Hörfunk, Fernsehen, Internet. 6. überarbeitete und erweiterte Auflage des „Gebrauchswörterbuch Fernsehen“. München: TR-Verlagsunion. (2000)

**Fahrmeir, Ludwig; Künstler, Rita; Pigeot, Iris und Tutz, Gerhard:** Statistik - Der Weg zur Datenanalyse. 3., verbesserte Auflage. Berlin, Heidelberg, New York: Springer-Verlag. (2001)

**Faller, Christof und Baumgarte, Frank:** Binaural Cue Coding - Part II: Schemes and Applications. (2003)  
zu beziehen unter <http://infoscience.epfl.ch/record/54881>

**Görne, Thomas:** Tontechnik. 2., aktualisierte Auflage. München: Carl Hanser Verlag (2008)

**Herre et al.:** The Reference Model Architecture for MPEG Spatial Audio Coding, AES Convention Paper 6447. 118. AES Convention Barcelona. (2005)  
zu beziehen unter [http://www.iis.fraunhofer.de/Images/AES6447\\_MPEG\\_Spatial\\_Audio\\_Reference\\_Model\\_Architecture\\_tcm97-67568.pdf](http://www.iis.fraunhofer.de/Images/AES6447_MPEG_Spatial_Audio_Reference_Model_Architecture_tcm97-67568.pdf)

**International Telecommunication Union (ITU) (Hrsg.):** RECOMMENDATION ITU-R BS.775-2 - Multichannel stereophonic sound system with and without accompanying picture. (2006)  
zu beziehen unter <http://www.itu.int/rec/R-REC-BS.775-2-200607-I/en>

**International Telecommunication Union (ITU) (Hrsg.):** RECOMMENDATION ITU-R BS.1534-1 - Method for the subjective assessment of intermediate quality level of coding systems. (2003)  
zu beziehen unter <http://www.itu.int/rec/R-REC-BS.1534/en>

**Lee, Jack Y. B.:** Scalable Continuous Media Streaming Systems - Architecture, Design, Analysis and Implementation. West Sussex: John Wiley & Sons Ltd. (2005)

**Lerch, Alexander:** Bitdatenreduktion. In: Handbuch der Audiotechnik. Hrsg. Prof. Dr. Stefan Weinzierl. Berlin - Heidelberg: Springer-Verlag (2008)

**Slavik, Karl M. und Weinzierl, Stefan:** Wiedergabeverfahren. In: Handbuch der Audiotechnik. Hrsg. Prof. Dr. Stefan Weinzierl. Berlin - Heidelberg: Springer-Verlag (2008)

## **Erklärung**

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe. Wörtlich oder sinngemäß übernommenes Gedankengut wurde im Text gekennzeichnet. Sämtliche verwendete Quellen sind im Literaturverzeichnis nachgewiesen.

Stuttgart den 03.11.2009

---

Gerhard Wicho