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

Inhaltsverzeichnis

Abstract .............................................................................................................. 2
Inhaltsverzeichnis .......................................................................................... 3
Abbildungsverzeichnis ..................................................................................... 5
1. Einleitung ..................................................................................................... 6
2. Begriffserläuterungen ................................................................................ 7
   2.1 Mehrkanalton ........................................................................................ 7
   2.2 Streaming .............................................................................................. 8
   2.3 Kodierverfahren und Bitratenreduktion ................................................. 9
3. Bandbreitenbedarf und Audiokodierverfahren ........................................... 10
   3.1 Datenraten bei der digitalen Informationsübertragung im Hörfunkbe-
      reich 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 Audiokodie-
         rung ..................................................................................................... 17
4. Grundlagen der Streamingtechnologie ...................................................... 18
   4.1 Das Client-Server Modell ..................................................................... 18
   4.2 Anforderungen an Streamingsysteme ................................................... 18
   4.3 Streamingfähige Protokolle .................................................................. 19
5. Erläuterung der MUSHRA Methode ....................................................... 21
   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 statisti-  
schen Analyse ..............................................................................................25  

6. Durchführung der Hörtests .......................................................................26  
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 Trainingssession ......................................................................................35  
6.8 Durchführungszeitraum und Dauer der Tests .........................................35  

7. Auswertung und grafische Aufbereitung ..................................................36  
7.1 Boxplot ..................................................................................................36  
7.2 Darstellung der Mittelwerte und Konfidenzintervalle nach Kodierverfah-  
ren 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  

8. Fazit ............................................................................................................58  
Literaturverzeichnis .......................................................................................59  
Erklärung ........................................................................................................60
Abbildungsverzeichnis

Abb. 1 Aufstellung 3/2 Mehrkanalsystem .................................................. 8
    aus ITU-R B.S. 775-2 (siehe Literaturverzeichnis)

Abb. 2 Prizipielles 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 Schneideraum 5) im Funkhaus München

Abb. 5 Blick auf Arbeitsplatz und Frontlautsprecher DSR5 ......................... 34
    Foto mit Blick über den Sweetspot und den Arbeitsplatz auf die Front-
    lautsprecher im DSR5
1. Einleitung


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.


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 Hörrer 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.¹

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-

¹ Vgl. ITU-R B.S. 775-2 S. 2 ff

**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° bzw. 48°; daraus ergeben sich die Abstände von 3 x H zu Bildschirm 1 (HDTV Referenz Abstand) und 2 x H zu Bildschirm 2, wobei H die Höhe des Bildschirmes beschreibt.

---

**2.2 Streaming**


\(^2\) 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.\(^3\)

### 2.3 Kodierverfahren und Bitratenreduktion


---

\(^3\) Vgl. Bayerischer Rundfunk (2000) S. 265

\(^4\) Vgl. Lerch (2008) S. 849 f
3. Bandbreitenbedarf und Audiokodierverfahren


3.1 Datenraten bei der digitalen Informationsübertragung im Hörfunkbereich des Bayerischen Rundfunks


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:

<table>
<thead>
<tr>
<th>Programm</th>
<th>Datenrate in kbps</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayern 1</td>
<td>320</td>
<td>Musicam</td>
</tr>
<tr>
<td>Bayern 2</td>
<td>320</td>
<td>Musicam</td>
</tr>
<tr>
<td>Bayern 3</td>
<td>320</td>
<td>Musicam</td>
</tr>
</tbody>
</table>
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.\footnote{Online: http://www.br-online.de/unternehmen/technik/rundfunktechnik-DID120066073148/rundfunktechnik-radio-satellit-ID671202493936395040.xml (Stand: 01.09.2009)}

### 3.1.2 DVB-C

Die Einspeisung ins Kabelnetz ist regional unterschiedlich und hängt vom jeweiligen Kabelnetzbetreiber ab. Über Programmvielfalt und Qualität lassen sich daher keine allgemein gültigen 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:

<table>
<thead>
<tr>
<th>Programm</th>
<th>Datenrate in kbps</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayern 2 plus (größtenteils identisch mit dem Programm von Bayern 2)</td>
<td>128</td>
<td>Musicam</td>
</tr>
<tr>
<td>BR Klassik</td>
<td>192</td>
<td>Musicam</td>
</tr>
<tr>
<td>B 5 plus (mono)</td>
<td>96</td>
<td>Musicam</td>
</tr>
<tr>
<td>Bayern plus</td>
<td>128</td>
<td>Musicam</td>
</tr>
<tr>
<td>on3</td>
<td>128</td>
<td>Musicam</td>
</tr>
<tr>
<td>BR Verkehr (mono)</td>
<td>64</td>
<td>Musicam</td>
</tr>
</tbody>
</table>

Im Digitalradio variiert die Datenrate mehr als bei DVB-S. Da aus frequenztechnischen Grün-
den im Digitalradio die Programme nicht separat übertragen werden, sondern in Programm-
paketen (auch Ensembles genannt) mit einer festgelegten Bandbreite zusammengefasst wer-
den, 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 Qua-
lität für die Vermittlung der Information ausreicht.

In München wird zudem ein regional empfangbares Testensemble des Bayerischen Rund-
funks ausgesendet, in dem Bayern 1, Bayern 2 plus und Bayern 3 bei einer Datenrate von
192 kbps enthalten sind.\footnote{Online: \url{http://www.ukwtv.de/sender-tabelle/index.html} (Stand: 01.09.2009)}

Zu Testzwecken wurde am 28. Dezember 2008 einen Tag lang vorproduziertes Mehrkan-
altonmaterial über das Digitalradio verbreitet. Ein Datenformat, welches für Mehrkanal-
übertragung via Digitalradio in Zukunft in Frage kommen könnte und an diesem Tag auch
ausgestrahlt wurde, ist MPEG Surround. In dem Datenstrom wurde ein Stereosignal weiter-
gegeben sowie Zusatzinformationen, die aus der ursprünglichen Mehrkanalmischung errech-
net 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 je-
welts vier Varianten aller empfangbaren Programme (Bayern 1, Bayern 2, Bayern 3,
BR Klassik, B5 aktuell, B5 plus, Bayern plus, on3) auf seiner Homepage \url{www.br-online.de}
zur Verfügung:

<table>
<thead>
<tr>
<th>Programm</th>
<th>Datenrate in kbps</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>alle gestreamten Programme sind in vier Varianten verfügbar</td>
<td>56</td>
<td>MP3</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>Windows Media Audio</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>128</td>
<td></td>
</tr>
</tbody>
</table>
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.\footnote{Online: http://www.br-online.de/unternehmen/technik/rundfunktechnik-DID120066073148/rundfunktechnik-radio-livestreams-ID67120249394689045.xml (Stand: 01.09.2009)}

### 3.2 Ansätze von Audiokodierverfahren zur Bitratenreduktion

#### 3.2.1 Verlustlose Kodierung


#### 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
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.


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 Frequenzauflosung angestrebt, bei länger klingenden Signalen können diese bei gleichzeitiger Erhöhung der Frequenzauflosung 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

3.2.3 Räumliche Parametrische Kodierung

Für die Kodierung von mehrkanaligem Audiomaterial gibt es nun weitere, allerdings verlustbehafte 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

\[ Vgl. \text{Lerch (2008) S. 857 ff} \]
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 wiederzustellen.\(^9\)

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.\(^{10}\)

---

9 Vgl. Faller und Baumgarte (2003) S. 520 ff
3.2.4 Wahrnehmbare Qualitätsverluste bei verlustbehafteter Audiokodierung


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 Rauhigkeit.

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.\[11\]

\[11\] 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.

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.\textsuperscript{12}

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 vonstattengehen. 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 Latenzeiten schlecht vorherzusagen. Auch der Deko-

\textsuperscript{12} Vgl. Lee (2005) S. 3 ff
dieser kann aufgrund von parallel laufenden Prozessen auf der Clientseite unterschiedliche Verarbeitungszeiten aufweisen.


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 finanzieltem Aufwand verbunden, könnte aber auch im Falle von Überlastungen genutzt werden.\textsuperscript{13}

\section*{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.


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

\textsuperscript{13} Vgl. Lee (2005) S. 11 ff


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.¹⁴

¹⁴ Vgl. Lee (2005) S. 111 ff
5. Erläuterung der MUSHRA Methode


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.\textsuperscript{15} 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.\textsuperscript{15}

\textsuperscript{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 unberechtigte Verfälschungen vermieden.\textsuperscript{16}

### 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.\textsuperscript{17}

### 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.\textsuperscript{18}

\begin{itemize}
  \item \textsuperscript{16} Vgl. ITU-R BS.1534-1 S. 3
  \item \textsuperscript{17} Vgl. ITU-R BS.1534-1 S. 6
  \item \textsuperscript{18} Vgl. ITU-R BS.1534-1 S. 5f
\end{itemize}
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.\(^{19}\)
Um ein aussagekräftiges Ergebnis zu erhalten und entsprechende Schlüsse aus den Tests ziehen zu können, reichen 20 Testpersonen aus.\(^{20}\)

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.\(^{21}\)

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   \(u_{ij}\) 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.


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:

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

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.\(^\text{22}\)

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

\(^{22}\) 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 Ver-
teilung 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

\[ D_i = X_i - Y_i \quad i = 1, \ldots, n \]

wobei \( X_i \) Stichprobe eins und \( Y_i \) Stichprobe zwei darstellt

2. Man bildet die dazugehörigen betragsmäßigen Differenzen \(|D_1|, \ldots, |D_n|\)

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.\(^{23}\)

\(^{23}\) Vgl. Fahrmeir, Künstler, Pigeot, Tutz (2001) S. 428 f
6. Durchführung der Hörtests

6.1 Auswahl der Audiobeispiele


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:

<table>
<thead>
<tr>
<th>Titel</th>
<th>Gattung</th>
<th>Quelle</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maná - Vivir sin aire</td>
<td>Weltmusik</td>
<td>DVD-Audio</td>
<td>MLP 5.1, 24bit, 96kHz</td>
</tr>
<tr>
<td>Gustav Mahler - Allegro Maestoso aus 2. Sinfonie</td>
<td>Klassische Musik (großes Orchester)</td>
<td>DVD-Audio</td>
<td>MLP 5.1, 24bit, 96kHz</td>
</tr>
<tr>
<td>Crystal Method - Starting Over</td>
<td>Elektronische Popmusik</td>
<td>DVD-Audio</td>
<td>LPCM 5.1, 24bit, 48kHz</td>
</tr>
<tr>
<td>Manfred Leuchter - Ma Fi Minnak</td>
<td>Jazz</td>
<td>DVD-Video</td>
<td>DTS 5.0, 1536kbps</td>
</tr>
<tr>
<td>Jules Verne - Die Reise zum Mittelpunkt der Erde</td>
<td>Hörspiel</td>
<td>DVD-Video</td>
<td>DTS 5.0, 768kbps</td>
</tr>
<tr>
<td>Pjotr Iljitsch Tschaikowski - 4. Sinfonie</td>
<td>Klassische Musik (großes Orchester)</td>
<td>DVD-Video</td>
<td>DTS 5.0, 1536kbps</td>
</tr>
<tr>
<td>Johann Strauß - Wiener Blut</td>
<td>Operette (Orchester und Gesang)</td>
<td>DVD-Video</td>
<td>Dolby Digital 5.1, 448kbps</td>
</tr>
<tr>
<td>Giuseppe Verdi - Gefangenchor aus Nabucco</td>
<td>Oper (Orchester und Chor)</td>
<td>DVD-Video</td>
<td>Dolby Digital 5.1, 448kbps</td>
</tr>
</tbody>
</table>
In die Trainingssession hatte ich auf vier Materialien zurückgegriffen, die auch in der engeren Wahl für den eigentlichen Test gestanden waren:

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

### 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**


**6.2.4 Dolby Digital**


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 ord-
nen, sodann die Tonbeispiele in ihrer Länge festzulegen, entsprechend auszuschneiden und,
nachdem ich die Lautheit der Hörbeispiele untereinander angepasst hatte, als sechskanali-
ges Interleavedfile (wiederum 24 bit, 48 KHz) zu exportieren.

Nun konnte ich den Encodervorgang starten. Um die Windows Media Dateien zu erzeu-
gen, 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 Me-
dia Audio 10 Professional mit der Einstellung Two-pass encoding. Die daraus resultierenden
.wma-Dateien importierte ich wiederum in Nuendo, wo sie automatisch in sechskanalige in-
terleaved WAV-Dateien umgewandelt wurden und für den Test eingesetzt werden konnten.
Für die Erstellung der MP3 Surround Dateien steht auf der Seite des Fraunhofer Instituts 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 Firewireinterfaces 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 Enkodervorgang 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


6.5 Testsoftware


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

---

Abb. 3 Bedienoberfläche von STEP während eines Tests

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 Schneideraum 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 m² und einer mittleren Höhe von etwa 2,25 m.

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](image)

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
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 Klinkenkabel mit den Laut-
sprechern 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 ge-
messen.
6.7 Trainingssession

Während der Trainingssession wurden den Probanden die Funktionen der Testsoftware erläutert und sie hatten die Möglichkeit sich anhand der vier Testhörbeispiele zuhören und auf die Aufgabenstellung einzustellen. Das Audiointerface stellt einen Softwaremixer 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, dass am wenigsten vom Original abweicht. Zudem wurde auf Hidden Reference Anchor hingewiesen und deren Bedeutung für den Test erklärt. Während der Trainingssession war ich im Raum um eventuell aufkommende 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

7. Auswertung und grafische Aufbereitung


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.


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.\footnote{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.

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

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 dem entsprechend 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 Bewertungen 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 Bewertung 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 dazu schließen lässt, dass es sich nur um einzelne extreme Bewertungen 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

**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 Kodier- verfahren 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:

- k kenzeichnet den errechneten arithmetischen Mittelwert
- l bezeichnet den Bereich des Konfidenzintervalls in der zu 95% der tatsächliche Wert liegt

### 7.2.1 Alle Kodierverfahren über alle Hörbeispiele

![Diagramm der Bewertung der Kodierverfahren](image)


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

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.


7.2.2 Mittelwerte und Konfidenzintervalle aller Hörbeispiele separat für jedes Kodierverfahren


Mittelwerte aller Hörbeispiele für Kodierverfahren MP3 Surround 192 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.
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.
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 Sinfonieausschnitte 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.
Für MPEG Surround 192 kbps wurde bereits darauf hingewiesen, dass die Ergebnisse nur unter Vorbehalt mit den vorhergegangenen Verfahren zu vergleichen sind, denn die Probleme, die eine Beeinträchtigung der Qualität zur Folge haben könnten, sind maßgeblich davon beeinflusst, wie genau man die Stereofassung und die Mehrkanalfassung vor dem Encodervorgang übereinander legen konnte. So war es mir bei der Vorbereitung nicht möglich für die Beispiele Hörspiel (experimentell) und Weltmusik (Mundharmonika) die achtspurige Datei für den Encoder exakt anzulegen, da die Fassungen unterschiedlich geschnitten waren und in jedem Fall voneinander abweichen mussten. Bei der Operette wurde in der Mehrkanalfassung ein derart auffälliger Hall verwendet, der das Ergebnis des Encodervorgangs deutlich verfremdete. Diese Problematik schlägt sich nun auch in der Bewertung dieser drei Fälle nieder, die die schlechtesten Mittelwerte (im Poor Bereich) aufweisen. Dagegen finden sich die Barock- und Filmmusik sowie die Tschaikowsky Sinfonie im Bereich über 70 Punkte wieder. Bei der Weiterentwicklung des Verfahrens bzw. der Encodersoftware wäre es ratsam, an den beschriebenen Schwierigkeiten anzusetzen, denn mit dieser Thematik wird man vor allem bei vorproduziertem Material wahrscheinlich immer wieder konfrontiert.
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)

![Bewertungskurven für Kodierverfahren bei Mahler](image1.png)

Alle Kodierverfahren bei Hörbeispiel Mozart (Streichquartett)

![Bewertungskurven für Kodierverfahren bei Mozart](image2.png)
Alle Kodierverfahren bei Hörbeispiel Nabucco (Chor in Oper)

Alle Kodierverfahren bei Hörbeispiel Wiener Blut (Operette)
Alle Kodierverfahren für Hörbeispiel Tschaikowsky (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.

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 Stereooanwendungen auf, liefert allerdings selbst bei 192 kbps keine sehr stabile Qualität, wie die Analyse der Tests zeigte.


Literaturverzeichnis


Erklärung


Stuttgart den 03.11.2009

Gerhard Wicho