

Diplomarbeit

Downmixverfahren für MP3–Surround

von
Bastian Schick

Erich-Thienhaus-Institut der Hochschule für Musik Detmold
in Kooperation mit
Fraunhofer IIS, Erlangen

Betreuung HfM Detmold: Prof. Rainer Maillard (Erstgutachter)
Betreuung Fraunhofer IIS: Dipl.-Ing. Claus-Christian Spenger

Zeitraum: Mai - Juli 2004

hochschule *für musik*  detmold


Fraunhofer Institut
Integrierte Schaltungen

Inhaltsverzeichnis

1	Einleitung	3
2	Funktionsweise von MP3-Surround	4
2.1	Funktionsweise von BCC	4
2.2	Funktionsweise von MP3-Surround	8
3	Aufgabenstellung und Lösungsansatz	9
4	Ablauf	10
4.1	Hörtest	10
4.2	Auswertung	12
5	Qualitätsevaluation der Stereomischungen	13
5.1	Beschreibung der automatischen Stereomischung	13
5.2	Stereohörtest	14
6	Untersuchung für Surround	16
6.1	Beschreibung der dekodierten Surroundmischung	17
6.2	Pegel	17
6.3	Panorama	19
6.4	Hall	21
6.5	Mikrofonierung	23
6.6	Delay	24
6.7	Dynamics	25
6.8	Effekte	27
7	Zusammenfassung	29
	Erklärung	30

Abbildungsverzeichnis

1	Schematische Darstellung des BCC-Verfahrens	5
2	Aufbau eines MP3-Surround-Encoders	8
3	Benutzeroberfläche der Testsoftware „Wavswitch“	10
4	Signalfluss des Testaufbaus	11
5	Ergebnisse der Hörtests zu den Stereomischungen	15
6	Ergebnis des Hörtest zum Beispiel „Drums“, Stereo	15
7	Ergebnis des Hörtests zum Parameter „Pegel“	18
8	Positionierung der Schallquellen beim Beispiel „Bläser“	19
9	Ergebnis des Hörtests zum Parameter „Panorama“	20
10	Ergebnis des Hörtests zum Parameter „Hall“	22
11	Hörtest z. Parameter „Hall“: Median	23
12	Ergebnis des Hörtests zum Parameter „Mikrofonierung“	24
13	Ergebnis des Hörtests zum Parameter „Delay“	25
14	Ergebnis des Hörtests zum Parameter „Dynamics“	26
15	Ergebnis des Hörtests zum Parameter „Effekte“	28

1 Einleitung

ISO/MPEG-1/2-Layer 3, kurz MP3, ist eines der wohl bekanntesten Kodierverfahren für Audiosignale und hat in den letzten 12 Jahren eine weite Verbreitung gefunden. MP3 unterstützt Mono- und Stereo¹-Signale mit einer Samplingrate bis 48 kHz bei einer Bitrate bis zu 320 kBit/s. Damit war MP3 gut geeignet für Audio-Übertragung und -Speicherung mit geringer Datenrate, wie z.B. für Internetanwendungen (Download und Streaming) und mobile Applikationen (z.B. in Form von tragbaren MP3-Playern).

Mit der Verbreitung der DVD-Video und dem damit verbundenen vermehrten Aufkommen von Surround-Wiedergabesystemen ging auch ein steigendes Angebot an Mehrkanal-Audiomaterial einher. Abgesehen von der DVD-Audio und der SACD, welche mit unkomprimierten PCM- oder DSD-Ton arbeiten, muss das Mehrkanal-Audiomaterial für die Übertragung oder Speicherung datenkomprimiert werden. Für die DVD-Video stehen dafür die Formate AC-3 (384-465 kBit/s) und DTS (754 & 1509 kBit/s) zur Verfügung. Für Anwendungen, welche eine stärkere Kompression erfordern, wurde ISO/MPEG-2/4 Advanced Audio Coding (AAC) entwickelt, welches mit Datenraten von 256-320 kBit/s für 5-Kanal-Audio arbeitet.

Die oben genannten Datenraten sind jedoch für viele Anwendungen, wie z. B. Internet-Live-Streaming, immer noch relativ hoch und die kodierten Mehrkanal-Audiosignale nicht stereokompatibel, sondern nur auf entsprechenden Mehrkanalwiedergabesystemen spielbar. Aufgrund dieser Einschränkungen der herkömmlichen Kodierverfahren wurde MP3-Surround entwickelt: ein Kodierverfahren für Mehrkanal-Audiosignale, welches einerseits eine geringe Bandbreite aufweist und andererseits noch von herkömmlichen Stereowiedergabesystemen unterstützt wird.

MP3-Surround gehört zu einer neuen Generation von Kodierverfahren für Mehrkanal-Audiosignale. Basierend auf dem „Enhanced Binaural Cue Coding“-Verfahren (EBCC) wird dabei im Gegensatz zur diskreten Übertragung von 5.1-Material eine Stereomischung des Mehrkanalsignals zum Empfänger übertragen zusammen mit einer kompakten Seiteninformation, welche die gehörrichtige Expansion dieses Stereosignals in ein möglichst originalgetreues Mehrkanal-Klangbild ermöglicht. Durch die Übertragung von 2 statt 5 (oder mehr) Kanälen ergibt sich bereits eine Datenreduktion von bis zu 60 %. Die Seiteninformation weist dabei eine sehr geringe Datenrate auf (ca. 16 kBit/s bei 6-kanaligem Audiomaterial), sodass die Gesamtbitrate fast der eines kodierten Stereosignals entspricht.

Gleichzeitig ist dadurch eine Abwärtskompatibilität zu schon vorhandenen MP3-Stereowiedergabesystemen gewährleistet. Anwendungsgebiete für MP3-Surround sind z.B. Musik-Download-Dienste, Internet-Streaming (Internet-Radio), Digitaler Rundfunk und Audio für Computerspiele.

Die Stereomischung wird vom MP3-Surround-Encoder mittels eines automatischen Downmix-Algorithmus erstellt, welcher sein Downmix-Verhalten an das Audiomaterial anpasst [1]. Die qualitative Beurteilung dieses automatischen Downmix-Algorithmus stellt einen Teil der vorliegenden Untersuchung dar.

In vielen Fällen wird von Surround-Material schon während der Produktion eine Stereomischung hergestellt, um ein Maximum an künstlerischer Freiheit zu erhalten². Die dabei zur Verfügung

¹Der Begriff „Stereo“ bezieht sich in dieser Arbeit auf Zweikanal-Stereophonie.

²Für solche von Hand erstellten Abmischungen wird im weiteren Verlauf der Arbeit der Begriff „manuell“ verwendet.

stehenden Parameter reichen von einfachen Pegeländerungen bis hin zur Verwendung unterschiedlicher Mikrofonanordnungen für die Stereomischung. MP3-Surround bietet die Möglichkeit statt des vom MP3-Surround-Enkoders ertellten Stereomix auch andere (z.B. manuelle) Stereomischungen zu verwenden. Da sich die Seiteninformation jedoch auf den automatisch generierten Downmix bezieht, können bei der Verwendung alternativer Stereomischungen beim Dekodieren abweichende Ergebnisse auftreten. Die Auswirkungen unterschiedlicher Parameteränderungen in der Stereomischung auf die dekodierte Mehrkanalfassung werden im zweiten Teil der vorliegenden Untersuchung behandelt.

Die Gliederung dieser Arbeit gestaltet sich wie folgt: Kapitel 2 gibt eine Übersicht über die Funktionsweise sowohl des EBCC- als auch des MP3-Surround-Verfahrens. Das darauffolgende Kapitel 4 beschreibt den Aufbau und Ablauf der Untersuchung, der Hörtestreihen sowie die statistische Auswertung. Die Kapitel 5 und 6 behandeln die Ergebnisse der oben genannten Untersuchungen. Das abschließende Kapitel 7 fasst die Ergebnisse kurz zusammen.

2 Funktionsweise von MP3-Surround

MP3-Surround basiert auf dem „Enhanced Binaural Cue Coding“-Verfahren, im weiteren Verlauf dieser Arbeit kurz EBCC genannt, welches eine Erweiterung des „Binaural Cue Coding“-Verfahrens (BCC) ist. BCC wurde von Agere Systems, Allentown, USA entwickelt und ist ein Kodierverfahren, bei dem eine Trennung zwischen den für die räumliche Wahrnehmung notwendigen Informationen und der eigentlichen Audioinformation vorgenommen wird. BCC stellt somit Mehrkanalsignale durch ein Mono-Audiosignal und BCC-Parameter³ dar [5].

Diese Technik ist vergleichbar mit dem „Intensity stereo coding“-Verfahren (ISC), welches z.B. in MP3 oder AAC verwendet wird. Bei ISC werden die zwei Kanäle eines Stereosignals in jedem Subband durch ein Summensignal und eine Winkelangabe (Azimuth) ersetzt. Im Dekoder wird anhand des Azimuth durch Intensitätsspannung die ursprüngliche Position wieder hergestellt. Das Signal wird dabei in 1024 Subbänder zerlegt, welche wiederum durch 50 Skalenfaktorbander geteilt werden. Pro Skalenfaktorband wird ein Azimuth übertragen. Durch diese Auflösung im Frequenzbereich treten jedoch starke Artefakte auf, wendet man ISC auf die gesamte Bandbreite oder auf Audiosignale mit einer großen Räumlichkeit an. Diese Auflösung ist vom Programmkern des Codecs vorgegeben [8]. Bei BCC ist die Transformation vom Zeit- in den Frequenzbereich unabhängig vom Kodierverfahren selbst, sodass die Einschränkungen von ISC bei BCC durch Optimierung der Auflösung umgangen werden können.

In Kapitel 2.1 soll das BCC-Verfahren genauer beschrieben werden. Das Folgekapitel 2.2 geht auf die technischen Aspekte dieses Verfahrens bei MP3-Surround ein.

2.1 Funktionsweise von BCC

Es gibt zwei verschiedene Arten von BCC, Typ I, *BCC for flexible rendering*, und Typ II, *BCC for natural rendering*. Beim *flexible rendering* werden mehrere voneinander unabhängige Monosignale

³Die zur räumlichen Wahrnehmung relevanten Informationen werden in der englischsprachigen Literatur „spatial cues“ genannt. Dieser Begriff wird unübersetzt auch in dieser Arbeit verwendet. Die „spatial cues“ werden in Kapitel 2.1 genauer erklärt. Die durch das BCC-Verfahren errechneten „spatial cues“ werden als BCC-Parameter bezeichnet

kodiert und erst beim Dekodiervorgang virtuell räumlich angeordnet. Die „spatial cues“ werden in diesem Fall künstlich erzeugt. Dieses Verfahren ist für Anwendungen wie Teleconferencing gedacht. Für das En- und Dekodieren von Mehrkanalsignalen, wie es bei MP3-Surround der Fall, ist der Typ II, *BCC for natural rendering*, vorgesehen. Bei diesem Verfahren werden die „spatial cues“ aus dem gegebenen Material gewonnen, um so beim Dekodieren die ursprüngliche räumliche Darstellung des Mehrkanalsignals zu erhalten. Im folgenden wird lediglich die Funktionsweise von BCC Typ II betrachtet.

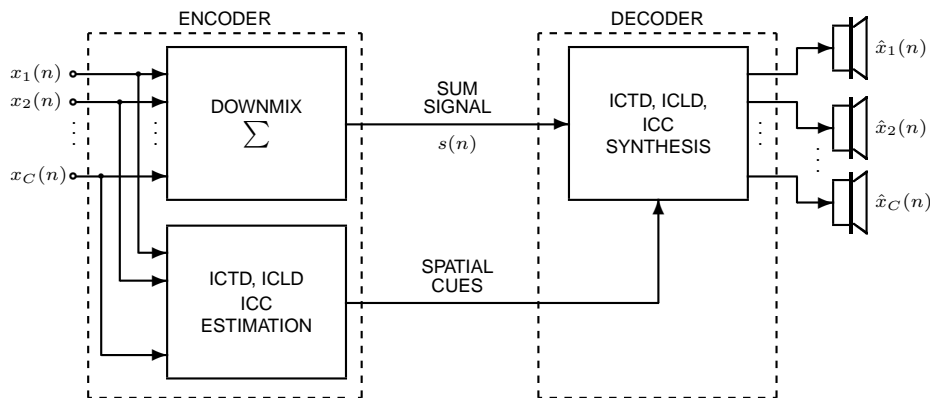


Abbildung 1: Schematische Darstellung des BCC-Verfahrens

Abbildung 1 zeigt den generellen Aufbau des BCC-Verfahrens. Aus den Eingangskanälen wird zuerst eine Monomischung erstellt, welche als Summensignal zum Empfänger übertragen wird. Gleichzeitig werden aus den Eingangskanälen die BCC-Parameter, auf die im folgenden noch genauer eingegangen werden soll, ermittelt. Diese Parameter werden wiederum als Seiteninformationen zum Dekoder übertragen, welcher aus dem Summensignal und den Seiteninformationen wieder das Mehrkanalsignal erstellt. Die Datenrate der Seiteninformationen bewegt sich in der Größenordnung von ca. 16 kB/s bei 6-Kanal-Audio und ist somit sehr gering.

Die Seiteninformationen beziehen sich auf die wichtigsten aus der Psychoakustik bekannten Größen für das räumliche Hören: die *interauralen Pegel-* (ILD) und *interauralen Zeitdifferenzen* (ITD) sowie die *interaurale Korrelation* (IC)⁴ an den Trommelfellen [9]. Für die Richtung einer Schallquelle sind dabei die ILD und ITD verantwortlich, die IC für die Breite einer Schallquelle.

Diese Größen sind jedoch schwer voraussehbar, da die akustische Übertragungsfunktion (ATF)⁵ von der Schallquelle zu den Ohren von verschiedenen Parametern, wie z. B. dem Wiedergabesystem und -raum abhängig ist. Da man allerdings davon ausgehen kann, dass bei Lautsprecherwiedergabe die Auswirkungen der ATF auf die dekodierte Fassung die gleichen wie auf das Originalsignal sind, werden für BCC lediglich die „spatial cues“ in der Übertragungskette zwischen den einzelnen Lautsprecherkanälen analysiert. Zu diesem Zweck wurden die Begriffe *Inter-channel level differences* (ICLD), *Inter-channel time differences* (ICTD) und *Inter-channel correlation* (ICC) eingeführt [2].

Für das Richtungshören von tiefen Frequenzen sind die interauralen Zeitdifferenzen (ITD) wichtig, bei hohen Frequenzen die interauralen Pegeldifferenzen (ILD). Die Grenzfrequenz liegt bei

⁴Die Abkürzungen beziehen sich auf die englischen Begriffe *interaural level differences* (ILD), *interaural time differences* (ITD) und *interaural correlation* (IC)

⁵ATF = acoustical transfer function

etwa 1,5 kHz. Versuche bei Lautsprecherwiedergabe im Freifeld haben ergeben, dass die ICLD am Ohr des Hörers wieder in ITD umgewandelt werden. Für Lautsprecherwiedergabe, wie sie bei MP3-Surround der Fall ist, kann man also die ICTD vernachlässigen⁶. In der Studioteknik ist diese Vorgehensweise ebenfalls weit verbreitet. So wird im Großteil der z. Zt. verfügbaren Mischpulte die Positionierung von Schallquellen auch für tiefe Frequenzen unterhalb von 1,5 kHz mittels Intensitätspannung realisiert. Für die empfundene Breite und Diffusität einer Schallquelle sind ist die ICC verantwortlich [2] [4].

Für die Analyse werden zunächst die Audiosignale in den Frequenzbereich transformiert. Dabei kommt eine *Fast Fourier Transformation (FFT)* zum Einsatz, da diese eine geringe Komplexität aufweist⁷. Nach der Transformation wird die gewonnene spektrale Darstellung in B nicht überlappende Partitionen mit dem Index b zerlegt. Jede Partition hat dabei eine Bandbreite, welche in etwa einer doppelten *Equivalent Rectangular Bandwidth (ERB)* entspricht. Die genauen Grenzen der Frequenzbänder zeigt Tabelle 1.

215 Hz	323 Hz	452 Hz	603 Hz
797 Hz	1034 Hz	1335 Hz	1723 Hz
2175 Hz	2756 Hz	3467 Hz	4350 Hz
5448 Hz	6826 Hz	8506 Hz	10594 Hz
13200 Hz	16430 Hz	20413 Hz	22050 Hz

Tabelle 1: Frequenzbandgrenzen

Für jede Partition b werden die Pegeldifferenzen ΔL_{ib} und Zeitdifferenzen τ_{ib} zwischen den einzelnen Kanälen berechnet. Dabei wird Kanal 1 als Referenz betrachtet. Die Inter-channel correlation Γ_b bestimmt die Korrelation aller Kanäle im Frequenzband b .

Für die Bestimmung der ICLD wird zunächst die Energie P in jedem Kanal $1 \leq c \leq C$ für jede Partition $1 \leq b \leq B$ ermittelt,

$$P_b^c = \sum_{n=A_{b-1}}^{A_b-1} |S_n^c|^2 \quad (1)$$

wobei S_n^c die Spektrallinien von Audiokanal c darstellen. Die bestimmten ICLD in dB zwischen Kanal c und der Referenz Kanal 1 für die Partition b lautet dann wie folgt:

$$\Delta L_{c-1,b} = 10 \log_{10} \left(\frac{P_b^c}{P_b^1} \right). \quad (2)$$

Für Frequenzen unterhalb von 1,5 kHz werden die ICTD durch Mittelung der Phasenverzögerung zwischen jedem Kanal und dem Referenzkanal pro Partition gewonnen. Die Formel für die ICTD lautet:

$$\bar{\tau} = \arg \max_d \{ \bar{\Phi}_{1c}(d) \} \quad (3)$$

mit der Kreuzkorrelationsfunktion Φ gemäß Gleichung (21) aus [4].

⁶Bei Kopfhörerwiedergabe sind die ICTD nicht vernachlässigbar. Dieser Fall ist jedoch für MP3-Surround irrelevant

⁷Das BCC-Verfahren bietet auch die Möglichkeit, das Audiosignal mittels einer *Cochlear Filter Bank (CFB)* kombiniert mit einem *Inner Hair Cell*-Modell zu parametrisieren. Dieser Vorgang geht allerdings mit einer Komplexitätszunahme einher, welche unnötig mehr Rechenleistung erfordert [3]

Für die Bestimmung der ICC werden für jede Partition pro Zeiteinheit die zwei stärksten Eingangskanäle betrachtet. Diese Vorgehensweise basiert auf der Überlegung, dass in den meisten Fällen Schallquellen mittels der Amplitude zwischen zwei benachbarten Lautsprechern positioniert werden und die Diffusität dieser Quelle durch die Korrelation der beiden Kanäle bestimmt wird. Der endgültige Grad der Korrelation errechnet sich aus der Gleichung

$$|\Gamma_b(k)|^2 = \frac{\sum_{n=A_{b-1}}^{A_b-1} |\Gamma(n, k)|^2 \Phi_{ii}(n, k) \Phi_{jj}(n, k)}{\sum_{n=A_{b-1}}^{A_b-1} \Phi_{ii}(n, k) \Phi_{jj}(n, k)}, \quad (4)$$

mit der gleichen Kreuzkorrelationsfunktion wie oben.

Während des Dekodiervorgangs werden die Spektrallinien S_c des Summensignals mit den ermittelten „spatial cues“ modifiziert und mittels einer inversen FFT wieder in den Zeitbereich transformiert. Diese Modifikation erfolgt für jeden Kanal c gemäß

$$S_{c,n} = F_{c,n} G_{c,n} S_n, \quad (5)$$

wobei $F_{c,n}$ die Pegel- und $G_{c,n}$ die Phasenveränderungen beschreiben. Diese ermitteln sich wie folgt:

$$F_{c,n} = 10^{\Delta L_{c-1,b} + r_{c-1,n}/20} F_{1,n} \quad (6)$$

$$G_{c,n} = \exp\left(-j \frac{2\pi n(\tau_{c-1,b} - \tau)}{N}\right). \quad (7)$$

Die Faktoren $F_{c,n}$ für den Referenzkanal ($c = 1$) werden so berechnet, dass für jede Partition die Summe der Energie aller Kanäle mit der Energie des Summensignals gleich ist. Dadurch wird eine Lautheit erreicht, die weitgehend unabhängig von den Pegelunterschieden $\Delta L_{i,b}$ ist. Die Normalisation

$$F_{1,n} = \frac{1}{\sqrt{1 + \sum_{i=1}^{C-1} 10^{(\Delta L_{i,b} + r_{i,n})/10}}} \quad (8)$$

ist üblich für Intensitätspannung.

Die Phasenverzögerung für den Referenzkanal ($c = 1$) wird so berechnet, dass die größtmögliche absolute Verzögerung für jeden beliebigen Kanal in der entsprechenden Partition minimal ist. Die Formel hierfür lautet:

$$G_{1,n} = \exp\left(-j \frac{2\pi n \tau_b}{N}\right) \quad (9)$$

mit

$$\tau_b = \frac{(\max_{1 \leq i \leq C} \tau_{ib} + \min_{1 \leq i \leq C} \tau_{ib})}{2}. \quad (10)$$

2.2 Funktionsweise von MP3-Surround

Für MP3-Surround wurde das BCC-Verfahren dahingehend erweitert, dass statt einer Monomischung als Summensignal eine Stereomischung zum Einsatz kommt. Dementsprechend wurde auch der Algorithmus für den Einsatz mit einer Stereomischung optimiert. Dieses erweiterte BCC-Verfahren nennt sich „Enhanced Binaural Cue Coding“-Verfahren (EBCC) und ist eine Entwicklung des Fraunhofer IIS, Erlangen, Deutschland und Agere Systems, Allentown, USA.

Abbildung 2 zeigt den Aufbau eines MP3-Surround-Encoders für eine 5.0-Mischung. MP3-Surround ist jedoch keineswegs auf fünf Kanäle limitiert, sondern kann für weitere Eingangskanäle erweitert werden, inklusive eines (oder mehrerer) „Low frequency effects“(LFE)-Kanäle.

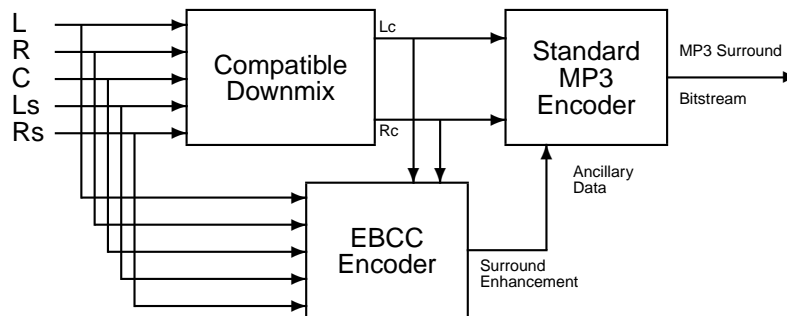


Abbildung 2: Aufbau eines MP3-Surround-Encoders

Der MP3-Surround-Encoder erstellt eine Stereoabmischung der Eingangskanäle und kodiert diese Stereomischung mittels eines herkömmlichen MP3-Stereo-Encoders. Desweiteren analysiert ein EBCC-Encoder die Eingangskanäle auf die „spatial cues“ hin und erstellt die Seiteninformationen. Diese werden dann in den MP3-Stereo-Datenstrom eingebettet. MP3-Stereo-Dekoder ignorieren diese Zusatzinformationen im Datenstrom und spielen ganz normal die Stereomischung ab, MP3-Surround fähige Systeme können die Zusatzdaten erkennen und entsprechend die Stereomischung in die Mehrkanalfassung dekodieren.

Da in der Stereomischung an sich mehr Informationen enthalten sind als in einer Monomischung, ergibt sich eine Qualitätssteigerung bei EBCC. So bleiben z. B. Breite und Diffusität von Schallquellen weitestgehend erhalten, sodass diese beiden Parameter nicht ausschließlich durch die ICC dargestellt werden müssen. Auch die ICLD und die ICTD können so differenzierter bestimmt werden.

Die Verwendung einer Stereomischung hat zusätzlich noch den Vorteil, dass dadurch die Abwärtskompatibilität zu schon vorhandenen MP3-Stereosystemen gewährleistet ist. Die Stereomischung wird mit einem dynamischen Downmixverfahren erstellt. Dieses Downmixverfahren versucht unerwünschte Auslöschungen oder Verstärkungen, welche bei der Addition von Signalen mit korrelierten Anteilen entstehen können, zu vermeiden oder zu kompensieren.

Dazu wird im Frequenzbereich die Summe der Kanäle pro Partition gebildet⁸. Anschließend wird die Energie jeder Partition P_b des Summensignals als auch der Eingangskanäle $P_{c,b}$ ($1 \leq c \leq C$) gemäß Gleichung (1) berechnet. Dann wird jede Partition des Summensignals mit einem Verstärkungsfaktor

$$g_b = \sqrt{\frac{\sum_{c=1}^C P_{c,b}}{P_b}} \quad (11)$$

versehen, sodass die Energie im Summensignal für jede Partition der Gesamtenergie aller Eingangskanäle für diese Partition entspricht. Dadurch werden Auslöschungen und Verstärkungen weitestgehend vermieden [4]. Um zu verhindern, dass auf diese Weise Signale, die sich nahezu vollständig ausgelöscht haben, wieder auf eine hohe Energie hochskaliert werden, gibt es eine Begrenzung in der maximalen Verstärkung.

Neben der Verwendung des auf diese Weise automatisch generierten Downmixes besteht bei MP3-Surround die Möglichkeit, einen manuell erstellten Downmix zu verwenden.

3 Aufgabenstellung und Lösungsansatz

Bei der Verwendung alternativer Stereomischungen als Summensignal können beim Dekodieren abweichende Ergebnisse auftreten, da sich die EBCC-Seiteninformationen auf den automatisch generierten Downmix beziehen. Es stellt sich also die Frage, wie stark diese Abweichungen ausfallen, wenn z. B. eine Signalkomponente in der manuell erstellten Stereomischung nicht vorhanden ist oder sich an einer anderen Position als im automatisch erstellten Downmix befindet. Die Auswirkungen unterschiedlicher Parameter in der Stereomischung auf die dekodierte Surroundfassung wurden in der vorliegenden Arbeit untersucht.

Zu diesem Zweck wurden verschiedene Surroundmischungen erstellt und mittels eines MP3-Surround-Codecs en- und dekodiert. Zusätzlich wurden verschiedene Stereomischungen angefertigt, in denen gezielt einzelne Parameter verändert wurden. Diese modifizierten Stereomischungen wurden dann als Grundlage für den Dekodierprozess verwendet. Die dabei entstandenen Abweichungen wurden daraufhin mittels eines Hörversuchs bewertet. Den genauen Ablauf der Untersuchung beschreibt Kapitel 4.

Zusätzlich fand noch eine qualitative Bewertung des automatischen Downmixalgorithmus von MP3-Surround im Vergleich zu statischen und manuellen Downmixverfahren statt, ebenfalls in Form eines Hörversuchs.

Während des En- und Dekodierens wurde bei dieser Untersuchung auf das MP3-Stereokompressionsverfahren verzichtet. Es wurde also nur der mittels EBCC kodierte Teil des En- und Dekodierprozesses untersucht. An dieser Stelle sei erwähnt, dass sich MP3-Surround noch in der Entwicklung befindet. Die vorliegende Arbeit wurde mit einer Zwischenversion vom Mai 2004 erstellt.

Eine Schwierigkeit der Hörversuche bestand in der Tatsache, dass die Probanden zwischen künstlerischen Qualitätsunterschieden und technischen Artefakten zu unterscheiden hatten. Schließ-

⁸Für den linken Kanal werden die Eingangskanäle „Links vorne“, „Links hinten“ und „Center“ addiert, für den rechten Kanal die Eingangskanäle „Rechts vorne“, „Rechts hinten“ und „Center“

lich sollte nicht untersucht werden, ob eine Fassung ohne Schlagzeug besser als eine Fassung mit Schlagzeug ist, sondern wie weit dabei eventuell auftretende Störgeräusche die Qualität beeinträchtigen. Bei der Bewertung der Stereomischungen stellte sich das Problem, auf welche Referenz sich die Qualitätsbeurteilungen beziehen sollten. Letztendlich wurde die ursprüngliche Surroundmischung als Referenz gewählt, sodass die Probanden nun jedoch zwischen der Surroundmischung und den Stereomischungen abstrahieren mussten. Eine genaue Beschreibung der Hörversuche und die damit verbundenen Schwierigkeiten beschreibt Kapitel 4.1.

4 Ablauf

4.1 Hörtest

Die Hörtests wurden gemäß der ITU-Empfehlung BS.1534-1, „Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)“ ausgeführt. Das MUSHRA-Verfahren (MUSHRA: **M**U**l**t**i** **S**t**i**m**u**l**u**s **t**e**s**t **w**i**t**h **H**id**d**e**n** **R**e**f**e**r**e**n**c**e** **a**n**d** **A**n**c**h**o**r) ist ein Doppelblindtest, bei dem die Probanden jederzeit zwischen den einzelnen Stimuli umschalten können. Das unbearbeitete Original — im Falle dieser Untersuchung die unkodierte Surroundmischung — dient bei der Bewertung als Referenz und ist als einziges Beispiel bekannt. Zusätzlich wird die Referenz unter den zu beurteilenden Beispielen versteckt (Hidden reference). Desweiteren wird eine bei 3.5 kHz tiefpassgefilterte Version des Original, der sogenannte „Anchor“ dem Testfeld hinzugefügt. Die „Hidden reference“ und der „Anchor“ sollen die Extremwerte der persönlichen Beurteilung der Probanden darstellen. Diese erfolgte pro Stimulus auf einer quasi-kontinuierlichen Skala, welche in 5 gleiche Abschnitte gemäß den Attributen „Bad“, „Poor“, „Fair“, „Good“ und „Excellent“ geteilt war, sodass für die Versuchspersonen eine grafische Wertung möglich war [13].

Der MUSHRA-Test wurde mittels der Software „Wavswitch“ des Fraunhofer IIS, Erlangen, durchgeführt. Die Oberfläche dieser Software zeigt Abbildung 3.

```

Press:

<1> for Reference ( 0 dB)
<2> for Testitem ( 0 dB) [+-----+-----#-----+-----+-----]
<3> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----]
<4> for Testitem ( 0 dB) [+-----+-----+-----+-----#-----]
<5> for Testitem ( 0 dB) [#-----+-----+-----+-----+-----]
<6> for Testitem ( 0 dB) [+-----+-----+-----+-----+-----#]
<7> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----]
<8> for Testitem ( 0 dB) [+-----+-----+-----+-----#-----]
<9> for Testitem ( 0 dB) [+-----+-----#-----+-----+-----]
                                bad      poor      fair      good      excellent

<a> to set loop start (<A> to reset): START
<b> to set loop end (<B> to reset): END

<+> to increase rating of playing item
<-> to decrease rating of playing item

<space> for stop/restart

<shift + q> for next item

6.27 s

```

Abbildung 3: Benutzeroberfläche der Testsoftware „Wavswitch“⁹.

⁹Über die Tastatur konnten die Probanden die Kreuze auf der Skala verschieben und so ihre Wertung abgeben

Die Hörtests fanden am Erich-Thienhaus-Institut in Detmold statt. Dafür wurde eigens ein Abhörraum gemäß der MUSHRA-Spezifikation eingerichtet, welche sich nach der ITU-Empfehlung BS.1116, „Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems“ richtet [14].

Als Wiedergabesystem kamen fünf Lautsprecher vom Typ „Nautilus 802“ der Firma B&W sowie der „RMB 1095“-5-Kanal-Verstärker der Firma Rotel zum Einsatz. Als Wandler diente ein RME „Hammerfall DSP Multiface“.

Die Software lief auf einem Dell „Inspiron 8200“. Abbildung 4 zeigt den Signalfluss des Testaufbaus. Das Audio lag mit einer Samplingrate von 44,1 kHz bei einer Auflösung von 16 bit in 6-Kanal-Dateien vom Typ „wav“ vor. Da für diese Untersuchung der LFE-Kanal nicht verwendet wurde, enthielt dieser Kanal digital null.

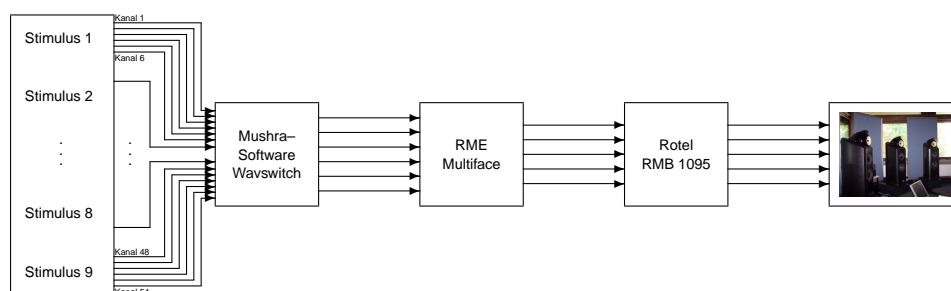


Abbildung 4: Signalfluss des Testaufbaus

Zu Beginn einer Testreihe gab es für jeden Probanden eine Einführung in die Thematik sowie die Erklärung der Testsoftware durch den Versuchsleiter. Darauf folgte eine Trainingsphase, um sich mit dem MP3-Surround-Kodierverfahren vertraut zu machen. Dabei wurden den Probanden sowohl das Original als auch das MP3-Surround-kodierte Material vorgespielt und die Möglichkeit gegeben, während der Wiedergabe zwischen den Hörbeispielen umzuschalten. Nach der Trainingsphase, deren Dauer die Probanden selbst bestimmen konnten, folgte die Bewertung der Surroundfassungen, danach die Bewertung der Stereomischungen.

Die Beurteilung gestaltete sich zum Teil als recht schwierig, da die Unterschiede zwischen den Surroundfassungen gelegentlich sehr groß waren. So mussten die Probanden z. B. bei der Untersuchung des Parameters „Pegel“ eine Mischung *ohne* Schlagzeug mit Mischungen *mit* Schlagzeug beurteilen (s. Kapitel 6.2, Seite 17). Aus diesem Grund wurden die Probanden gebeten, nur die generelle Audioqualität der Beispiele zu beurteilen. Dabei wurde ihnen nicht mitgeteilt, welcher Parameter in den zu Grunde liegenden Stereomischungen verändert wurde, um Fehler durch eine eventuelle Erwartungshaltung der Versuchspersonen zu vermeiden.

Für die Beurteilungen der Stereomischungen wurde als Referenz ebenfalls die ursprüngliche Surroundmischung gewählt. Die drei Stereomischungen lediglich untereinander vergleichen zu lassen, hätte zu einem wenig befriedigendem Ergebnis geführt, da ohne Bezugswert, auf den sich die Vergleiche hätten stützen können, die Beurteilungen stärker nach subjektiven Vorlieben ausgefallen wären. Die Probanden wurden entsprechend gebeten, die drei Stereomischungen nicht *mit* der Referenz zu vergleichen, sondern *in Bezug auf*.

Um die Bewertung mittels der MUSHRA-Software noch zu differenzieren, hatten die Probanden

die Möglichkeit, Auffälligkeiten und Beobachtungen auf einem Formular schriftlich festzuhalten.

Die Hörschaft setzte sich aus Lehrenden und Studierenden des Studiengangs Musikübertragung (Tonmeister) zusammen. Es wurden zwei Testreihen mit je 4 Surround- und 4 Stereobeispielen durchgeführt. Dabei haben bei der ersten Testreihe 18 Personen und bei der zweiten Testreihe 15 Personen teilgenommen. Bei der ersten Testreihe wurden nachträglich zwei Probanden von der Auswertung ausgeschlossen, da sie sowohl die versteckte Referenz als auch den Anchor inkorrekt eingeordnet hatten. Diese Fehler lassen sich zwar eher durch ein Missverständnis der Skala als durch mangelnde Hörfähigkeit erklären, ließen aber dementsprechend Zweifel an der Aussagekraft der Bewertung aufkommen. Beim zweiten Testdurchlauf wurde aus demselben Grund eine Person ausgeschlossen, sodass insgesamt 16, bzw. 14 Beurteilungen für die Auswertung verwendet werden konnten.

4.2 Auswertung

Die MUSHRA-Software erzeugte für jeden Probanden und jedes Testbeispiel eine Datei mit den bereits auf eine normierte Skala von 0 bis 100 angepassten Ergebnissen. Diese Dateien wurden von der Software IAQstat ausgewertet und grafisch aufbereitet.

Als Grundgröße liegt der Auswertung dabei das ungewogene arithmetische Mittel zu Grunde, welches gemäß der Formel

$$\bar{u}_{jk} = \frac{1}{N} \sum_{i=1}^N u_{ijk} \quad (12)$$

mit

$$\begin{aligned} u_{ijk} &= \text{Ergebnis von Proband } i \text{ für Hörbeispiel } j \text{ und Audiobeispiel } k \\ N &= \text{Anzahl der Probanden} \end{aligned}$$

berechnet wird.

Die Standardabweichung der Ergebnisse ergibt sich aus der Formel

$$S_{jk} = \sqrt{\sum_{i=1}^N \frac{(\bar{u}_{jk} - u_{ijk})^2}{(N-1)}} \quad (13)$$

und liefert als Standardabweichung der Stichprobe durch den Nenner $(N-1)$ die Standardabweichung einer übergeordneten Gesamteinheit. [11]

Aus diesen beiden Werten ergibt sich das 95%-Konfidenzintervall, welches sich aus

$$[\bar{u}_{jk} - \sigma_{jk}, \bar{u}_{jk} + \sigma_{jk}] \quad (14)$$

mit

$$\sigma_{jk} = t_{0,05} \frac{S_{jk}}{\sqrt{N}} \quad (15)$$

errechnet. Dabei kennzeichnet t den Wert der standardnormalverteilten Zufallsvariablen, an dem die Verteilungsfunktion der Standardnormalverteilung den Wert $1 - \frac{\alpha}{2}$ annimmt. Bei einem Konfidenzintervall von 95% entspricht α einem Wert von 0,05 und t einem Wert von 1,96. [12] [13]

Diesen Satz habe ich abgeschrieben

Zusätzlich wird in Kapitel 6.4 auf Seite 21 noch auf den Zentralwert, bzw. Median eingegangen. Dieser berechnet sich wie folgt:

$$\bar{x}_Z = x_{\frac{n+1}{2}}, \quad \text{bzw.} \quad \bar{x}_Z = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}). \quad (16)$$

5 Qualitätsevaluation der Stereomischungen

Ein großer Vorteil des MP3-Surround-Kodiervorgangs liegt in der Abwärtskompatibilität zu herkömmlichen Stereosystemen. Der Anwender kann je nach Abhörsituation, technischer Ausstattung und persönlicher Vorlieben zwischen der Stereo- und der Surroundfassung wechseln. Diese Tatsache ist jedoch nur so lange ein Vorteil, solange der Stereomix auch qualitativ gut ist. Wie weit der automatische Downmixalgorithmus von MP3-Surround dieser Forderung nachkommt, war eine Ausgangsfrage der vorliegenden Arbeit.

In diesem Kapitel wird zunächst eine qualitative Beschreibung des Downmixalgorithmus von MP3-Surround vorgenommen. Darauf folgt eine Besprechung der Ergebnisse der Hörtests zu den Stereomischungen.

5.1 Beschreibung der automatischen Stereomischung

Bei MP3-Surround werden für den automatischen Downmix jeweils die Frequenzbänder der Eingangskanäle addiert und durch Amplituden-Modifikation so angepasst, dass die Energie im Downmix gleich der Gesamtenergie der Eingangskanäle ist (siehe auch Kapitel 2.2). Dadurch hat der in MP3-Surround verwendete Downmix statischen Downmixverfahren gegenüber einen Vorteil, da durch Auslöschungen oder Verstärkungen verursachte Verzerrungen kompensiert werden. Außerdem wird die Qualität der Stereofassung dadurch vom Material unabhängig. Gerade bei Aufnahmen, welche wichtige Informationen, bzw. Direktschall auch in den Surroundkanälen enthalten, ist dies ein Vorteil. Bei den untersuchten Musikstücken betraf dies vor allem die Beispiele Jazz, Rock und die Bläseraufnahme, in denen einzelne Instrumente nur in den Surroundspuren enthalten waren. Diese Instrumente wurden im Gegensatz zu statischen Downmixverfahren gut in die Stereomischung integriert.

Bei zunehmender Lautstärke der Signalkomponenten in den Surroundkanälen besteht jedoch die Gefahr, dass diese Schallanteile im Downmix überzeichnet werden. Dies ist insofern ein Problem, als dass der Tonmeister, bzw. Produzent einzelne Signalkomponenten in den Surroundkanälen zur besseren Hörbarkeit deutlicher machen kann, da das Ohr für Schallereignisse von hinten weniger empfindlich ist. Diesen Fall konnte man bei den vorliegenden Musikstücken anhand der Triangel in der Jazzcombo oder der Klarinette bei der Bläseraufnahme beobachten. Dies ist auch für Auf-

nahmen relevant, deren Surroundkanäle hauptsächlich Diffusschall enthalten, da der Hallanteil in den Surroundkanälen für eine stärkere Umhüllung durchaus laut sein kann. Dadurch steigt auch der Hallanteil in der Stereomischung, vornehmlich die hohen Frequenzen, da diesen seltener ein Gegenstück im Direktschall entspricht. Diese Beobachtungen ergaben jedoch in keinem der Fälle ein unnatürliches Klangbild. Ob die Pegelunterschiede oder der stärkere Hallanteil erwünscht sind, ist im Einzelfall zu klären und stark von subjektiven Vorlieben abhängig.

Der Downmixalgorithmus stößt nur in dem Fall an seine Grenzen, wenn sich die Signale bei der Addition gegenseitig nahezu aufheben und dementsprechend die Amplituden-Modifikation wirkungslos bleibt. Dieses Problem mag bei der Beurteilung von Monokompatibilität von Stereosignalen Gewicht haben, weniger aber bei Surround, da die Wahrscheinlichkeit, dass alle fünf (oder mehr) Kanäle sich gegenseitig aufheben, äußerst gering ist. Dementsprechend trat dieser Fall auch bei keinem der Beispiele auf und kann kaum als Limitation angesehen werden.

Die Abbildung der Schallquellen im Stereodownmix stellt ebenfalls kein Problem für den MP3-Surround-Downmixalgorithmus dar. Die Frontabbildung bleibt erhalten und die hinteren Kanäle werden jeweils links und rechts hinzugefügt. Dabei werden auch Phantomschallquellen zwischen den Surroundkanälen sehr gut in der Stereomischung abgebildet. Bei Aufnahmen, die hauptsächlich diskrete Signale in den äußeren Kanälen enthalten, ergibt sich dadurch zwar eine Ballung der Informationen an den Rändern der Stereomischung, dieses wird jedoch meist als „offenes“ Klangbild bewertet, wie z.B. bei der Bläseraufnahme.

Durch dieses „offene“ Klangbild und den oben beschriebenen größeren Hallanteil weicht der automatische Downmix zwar von einer „optimalen“ Stereomischung in Einzelfällen ab, weist aber dafür eine größere Ähnlichkeit mit der Surroundmischung auf.

5.2 Stereohörtest

Um die Qualität der automatischen Stereomischung von MP3-Surround zu beurteilen, wurde ein Hörvergleich zwischen dieser automatischen Stereomischung, einem ITU-Downmix und einer manuell erstellten Stereomischung durchgeführt. Der manuelle Stereomix wurde dabei unabhängig vom Klangbild der Surroundmischung aus den Originalspuren der Aufnahme angefertigt.

Betrachtet man die Ergebnisse in Abbildung 5, stellt man fest, dass alle drei Varianten als gut bewertet wurden. Der automatische Downmix und der ITU-Downmix liegen dabei mit 69 bzw. 68,2 Punkten im Durchschnitt gleichauf. Erwartungsgemäß wurde der automatische Downmix dem ITU-Downmix bei den Aufnahmen vorgezogen, bei denen diskrete Informationen in den Surroundkanälen lagen, so z. B. bei der Bläseraufnahme, der Jazzcombo und der Rockmischung. Bei diesen Beispielen zeigten sich die im vorhergehenden Unterkapitel beschriebenen Vorteile des automatischen Downmixverfahrens von MP3-Surround. Die schlechteste Bewertung erfuhr der automatische Downmix mit 61,1 Punkten (was allerdings immer noch im Bereich „gut“ liegt) bei der Choraufnahme. In diesem Fall überwiegt der Hallanteil gegenüber dem Direktschall des Chores, sodass das Klangbild sehr räumlich wird.

Die manuellen Abmischungen sind mit 63,8 Punkten im Durchschnitt zwar ebenfalls „gut“ bewertet worden, aber dennoch etwas abgeschlagen gegenüber den automatischen Verfahren. Diese Tatsache ist jedoch sehr einfach zu erklären. Da die Stereomischungen unabhängig vom Klangbild der Surroundmischungen angefertigt wurden, sind sie aus Gründen des persönlichen Geschmacks generell etwas weniger hallig. Dadurch weisen sie jedoch eine größere Differenz zu der Original-

Beispiel: Stereo; 30 Hörer

Mittelwert und 95%-Konfidenzintervall

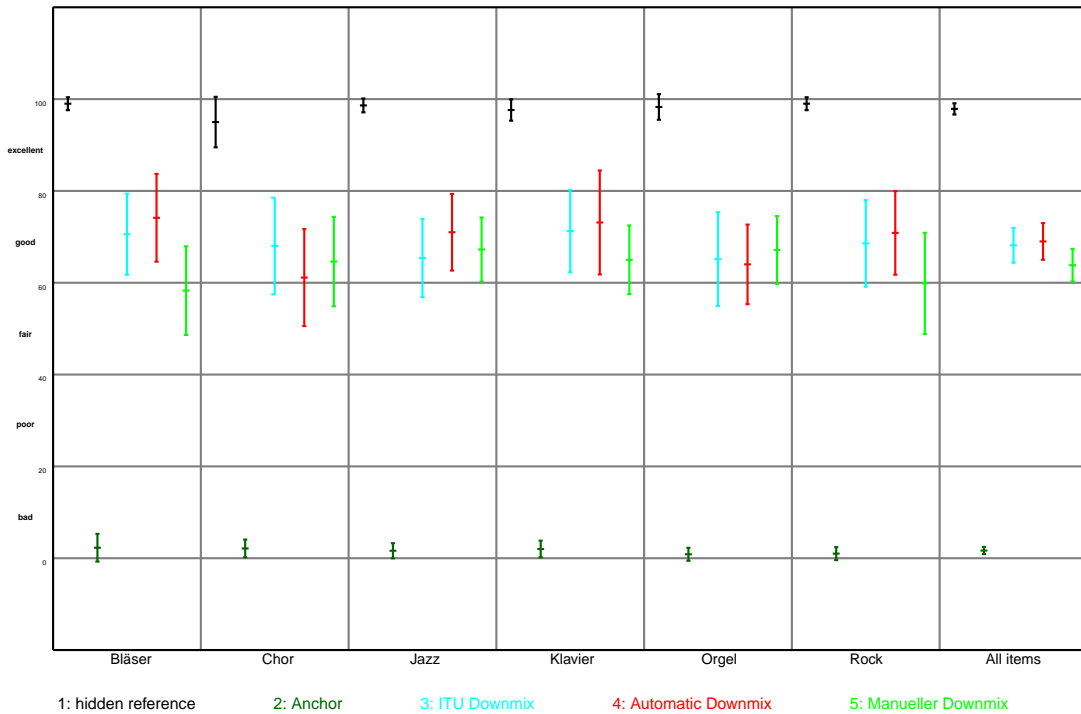


Abbildung 5: Ergebnisse der Hörtests zu den Stereomischungen

Surroundmischung auf. Da nun im Test die Ähnlichkeit zur Surroundmischung bewertet werden sollte, wurden die manuellen Mischungen geringfügig schlechter eingestuft. Sehr deutlich ist dies an den Bewertungen der Bläseraufnahme zu sehen, wo der Unterschied der Räumlichkeit sehr groß ist.

Interessant ist, dass bei der Orgelaufnahme alle Verfahren nahezu gleich bewertet worden sind. Bei dieser Aufnahme hatten alle Verfahren die gleichen Ausgangsbedingungen, da nur fünf Hauptmikrofone für die jeweiligen fünf Surroundkanäle verwendet wurden.

Die Beurteilung der Stereomischungen für das Beispiel „Drums“, wie Abbildung 6 zeigt, sollen an dieser Stelle gesondert betrachtet werden. Bei diesem Beispiel wurde die manuelle Stereomischung mit 37,9 Punkten im Durchschnitt wesentlich schlechter als die automatischen Downmixverfahren, deren Bewertungen bei 67,5 (ITU) und 68,9 (MP3-Surround) Punkten liegen, bewertet.

Generell war dieses Beispiel für jedes Verfahren mit Schwierigkeiten belastet. Wenn auch die Bewertungen der automatischen Downmixverfahren nicht darauf schließen lassen, so erfährt man doch aus den Kommentaren der Probanden, dass keines der Verfahren

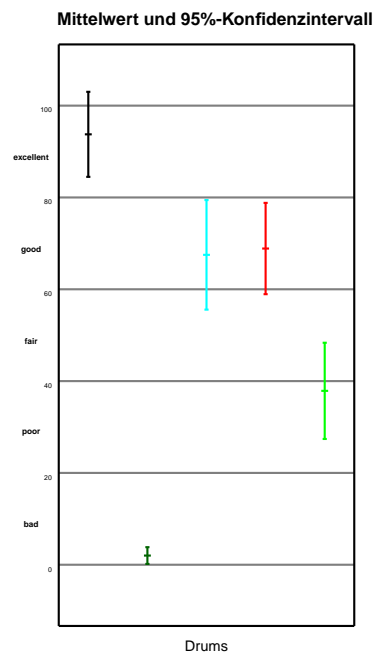


Abbildung 6: Ergebnis des Hörtest zum Beispiel „Drums“, Stereo

die Surroundmischung optimal umsetzen konnte. So wurde bei allen Verfahren eine „Phasigkeit“ und eine „unausgewogene Balance“ zwischen den einzelnen Kanälen bemängelt, was an dem durch frühe Reflexionen und Flatterechos geprägten Raum der Originalaufnahme liegen dürfte. Bei der manuellen Abmischung wurde zusätzlich kritisiert, dass die Abbildung sehr auf die linke Seite fokussiert sei.

Interessant ist, dass dieses Phänomen nur bei Lautsprecherwiedergabe auftritt. Bei persönlichen Beobachtungen bei Kopfhörerwiedergabe konnte ein gegenteiliger Effekt festgestellt werden: dort ist die manuelle Abmischung gleichmäßiger verteilt und die automatischen Verfahren linksbetont. In Ansätzen ist diese Differenz zwischen Lautsprecher- und Kopfhörerwiedergabe auch bei der Choraufnahme und dem Klavierbeispiel hörbar. Da als Referenz für die Hörtests Surroundmischungen genutzt wurden, war es nicht möglich auch die Qualität der Stereomischungen bei Kopfhörerwiedergabe zu testen. Da jedoch diese Form der Rezeption nicht unwichtig für die Verbreitung von MP3-Surround ist — man denke nur an die vielen tragbaren MP3-Player — sollte dieser Punkt mit weiteren Untersuchungen bedacht werden.

Neben der guten Qualität des automatischen Downmixverfahrens von MP3-Surround spielt natürlich auch die Geschwindigkeit des Verfahrens eine nicht unwesentliche Rolle. Da die Mischung in mehrfacher Echtzeit erstellt wird, bietet der Downmixalgorithmus eine schnelle, effiziente und dennoch qualitativ hochwertige Möglichkeit dar, Stereomischungen aus Surroundmaterial zu generieren. Damit stellt er auch für den Musikproduktionsprozess ein interessantes Werkzeug dar.

6 Untersuchung für Surround

Obwohl der Downmixalgorithmus von MP3-Surround von guter Qualität ist, kann es vorkommen, dass eine von der automatischen Stereomischung abweichende Fassung erwünscht ist. In vielen Fällen wird von Surround-Material schon während der Produktion eine Stereomischung hergestellt, um ein Maximum an künstlerischer Freiheit zu erhalten. MP3-Surround bietet in einem solchen Fall die Möglichkeit, eine eigene, manuell erstellte Stereomischung als Grundlage für den Dekodiervorgang zu verwenden, unter Beibehaltung der ursprünglichen EBCC-Parameter. Da diese sich jedoch auf die automatisch generierte Stereomischung beziehen, können Abweichungen der manuellen Stereomischung von der automatischen Stereomischung zu abweichenden Ergebnissen in der dekodierten Surroundfassung führen.

In diesem Kapitel sollen die Auswirkungen einzelner Parameterveränderungen in der manuellen Stereomischung auf die dekodierte Surroundfassung besprochen werden. Dabei wurden die Parameter „Pegel“, „Panorama“, „Hall“, „Mikrofonierung“, „Delay“, „Dynamics“ und „Effekte“ untersucht.

6.1 Beschreibung der dekodierten Surroundmischung

MP3-Surround ist kein verlustfreier Codec, woraus folgt, dass zwangsweise Unterschiede zu der Originalmischung auftreten müssen. Dabei unterscheiden sich die auftretenden Artefakte von den von anderen Kompressionsverfahren bekannten Artefakten.

Die dekodierte Fassung bei MP3-Surround weist häufig einen Pegelanstieg bei den tiefen Frequenzen gekoppelt mit einer Verunklarung derselben auf. Da dieser Bassanstieg auch schon in der Stereofassung auftritt, mag dort die Ursache für dieses Problem liegen. Desweiteren tritt eine Verengung der Abbildung ein. Dadurch wird das Klangbild schmal, und es geht Räumlichkeit verloren. Dieses Problem wurde allerdings inzwischen in den aktuelleren Versionen der MP3-Surround-Software mittels eines Diffusors behoben. Diese Versionen lagen jedoch leider nicht zum Zeitpunkt dieser Untersuchung vor. Außerdem können noch Rauigkeiten in einzelnen Frequenzbereichen auftreten, die man als „Blubbern“ charakterisieren kann. Wenn im weiteren Verlauf dieser Arbeit von Artefakten gesprochen wird, sind diese Rauigkeiten gemeint. Diese Artefakte werden im Gesamtklangbild weitestgehend verdeckt, sind jedoch sehr gut zu hören, wenn man sich die Kanäle separat anhört. Dies hat allerdings auch zur Folge, dass für MP3-Surround in noch viel stärkerem Maße eine Abhörsituation erforderlich ist, bei der alle Lautsprecher gleichwertig sind.

Diesen Einschränkungen unterlagen alle dekodierten Stereofassungen, sowohl die automatischen, als auch die manuellen mit und ohne Parameterveränderungen. Sie spielten für diese Untersuchung also nur insofern eine Rolle, dass sie durch einzelne Parameteränderungen entweder verstärkt oder kompensiert wurden.

6.2 Pegel

Einer der grundlegendsten Parameter ist zweifelsohne die Gesamtschalleistung einer Komponente in einer Abmischung. Dieser Parameter dürfte auch zu allererst zu Unterschieden zwischen einem automatisch und einem manuell erstellten Stereodownmix führen. Aus diesem Grund ist die Frage nach dem Verhalten des MP3-Surround-Dekoders bei Pegeländerungen im Stereodownmix von besonderer Bedeutung.

Um das Verhalten beim Dekodieren in Extremfällen zu untersuchen, wurden bei den manuellen Abmischungen gezielt Komponenten, d.h. Instrumente, entfernt. Als Beispiel wurde eine Aufnahme einer Jazzcombo mit der Besetzung Vibraphon, Klavier, Bass, Schlagzeug und Percussionsinstrumente (Agogo, Tamborin, Triangel) gewählt, da die einzelnen Komponenten akustisch weitgehend voneinander getrennt waren.

Die Gesamtschalleistung einer einzelnen Komponente wird bei MP3-Surround nicht separat gespeichert, lediglich die Gesamtschalleistung der einzelnen Frequenzbänder. Die Energie einer Komponente setzt sich also aus der Energie in den jeweiligen Frequenzbändern zusammen. Die daraus resultierenden ICLD werden jedoch nicht absolut, sondern relativ zu der Stereomischung angegeben. Eine Veränderung der Gesamtschalleistung einer Komponente im Stereodownmix zieht somit eine Veränderung der Gesamtschalleistung in der dekodierten Surroundmischung nach sich. Ist eine Komponente im Extremfall in der Stereomischung nicht vorhanden, taucht sie auch nicht in der dekodierten Surroundmischung auf.

Dabei erweist sich MP3-Surround als sehr robust gegenüber Pegeländerungen. Selbst wenn eine

Komponente in der Stereomischung komplett entfernt wird, sind die anderen Komponenten und ihre Abbildung weitestgehend intakt. Im Falle des vorliegenden Audiobeispiels war das an den Abmischungen ohne Bass und ohne Klavier und Vibraphon zu beobachten. Dementsprechend wurden diese Aufnahmen noch als „fair“ bewertet, wie man in Abbildung 7 sehen kann. Lediglich bei der Version ohne Bass traten ganz geringe Störgeräusche in Form von tieffrequenten Knacksern auf, die jedoch nicht als störend wahrgenommen wurden.

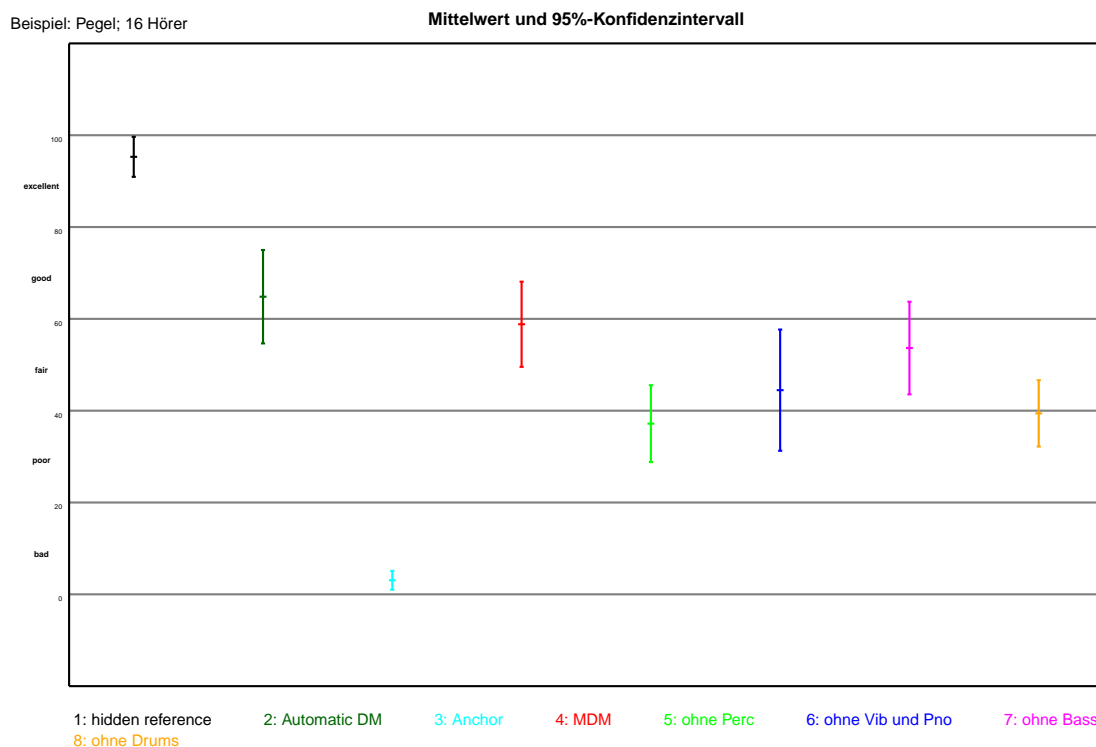


Abbildung 7: Ergebnis des Hörtests zum Parameter „Pegel“

Einschränkungen sind jedoch zu beachten, wenn eine Komponente entfernt wird, deren Frequenzanteile Frequenzbereiche von anderen Komponenten ursprünglich überdeckt, da die entsprechenden Frequenzanteile dann beim Dekodieren aus der anderen Komponente ersetzt werden.

Beim vorliegenden Beispiel der Jazzcombo trat dies ganz besonders bei den Percussionsinstrumenten hervor. Diese Instrumente, welche durch ihre Impulshaftigkeit ohnehin sehr viel Energie in mehreren Frequenzbändern enthalten, erwiesen sich als sehr anfällig beim Dekodieren. Durch Pegeländerungen überlagerten die Percussionsinstrumente nicht mehr vollständig andere Instrumente in den entsprechenden Frequenzbändern, vorzugsweise Schlagzeug, wodurch beim Dekodieren dann diese Instrumente den Percussionsinstrumenten „beigemischt“ wurden. Schon im manuellen Downmix, welcher noch ähnliche Pegelverhältnisse wie der automatische Downmix aufwies, machte sich dieser Effekt dadurch bemerkbar, dass die Percussionsinstrumente nicht mehr klar klangen, sondern von einem „Rauschteppich“ überlagert wurden. Ganz extrem fiel dies natürlich in dem Downmix auf, in welchem die Percussionsinstrumente ganz entfernt wurden. In der Surroundmischung ersetzte der Dekoder die fehlenden Signalkomponenten der Percussionsinstrumente mit Signalkomponenten des Schlagzeugs, was als sehr störend empfunden wurde. Die

Versionen ohne Percussionsinstrumente und ohne Schlagzeug wurden dementsprechend auch als durchschnittlich „poor“ bewertet.

Zusätzlich beeinflussen Pegeländerungen im Stereodownmix die Verdeckungseffekte im Gesamtklangbild, sodass plötzlich Frequenzbereiche hörbar sein können, die vorher von anderen Bereichen überdeckt wurden. Dadurch kann es zu einer Artefaktzunahme kommen. Gerade an der Klangqualität des Vibraphons im vorliegendem Beispiel kann man dies sehr gut hören.

Auf der anderen Seite kann man Pegeländerungen im manuellen Downmix auch nutzen, um andere Einschränkungen beim Dekodieren gezielt zu minimieren. So kann man z.B. die Energiezunahme im Bassbereich durch eine entsprechend bassärmere Stereomischung kompensieren.

6.3 Panorama

Ein weiterer wichtiger Parameter ist die Positionierung einer Schallquelle in einer Abmischung. Auch hier war die Ausgangsfrage, wie der Dekoder sich verhält, wenn die Position einer Schallquelle in der Stereomischung von der ursprünglichen Position im Surroundmix abweicht. Als Ausgangsmaterial wurde eine Bläserkammermusikaufnahme gewählt, bei der die Instrumente gemäß Abbildung 8a in der Surroundmischung platziert wurden.

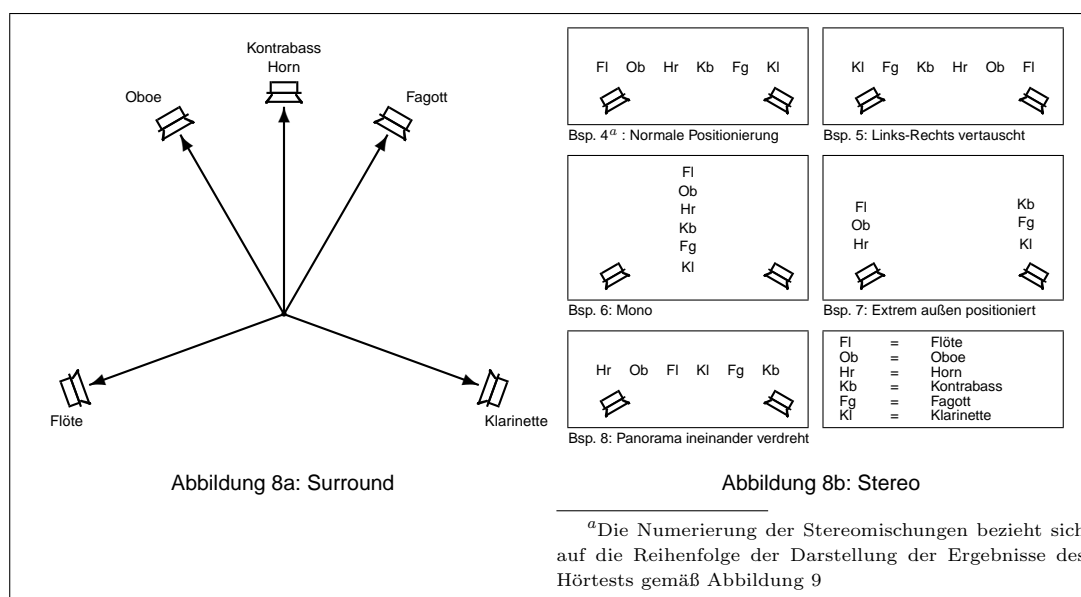


Abbildung 8: Positionierung der Schallquellen beim Beispiel „Bläser“

Diese Aufnahme enthielt für den MP3-Surround-Dekoder noch eine zusätzliche Schwierigkeit, da die Instrumente wegen Übersprechens der einzelnen Mikrofonspuren nicht absolut voneinander getrennt waren. Bei den manuellen Abmischungen wurden die Instrumente nach extremen Gesichtspunkten im Stereobild angeordnet, wie man in Abbildung 8b sehen kann.

Das relativ mittelmäßige Abschneiden der dekodierten manuellen Abmischungen in den Hörtests, wie man in Grafik 9 sehen kann, korreliert in diesem Beispiel mit den Ergebnissen der Hörtests zur Beurteilung der Stereoabmischungen (siehe Abbildung 5 auf Seite 15). Der Grund dafür dürf-

te in dem verringertem Räumlichkeitseindruck liegen, welcher schon bei der Stereoabmischung im Vergleich zur Mehrkanalreferenz zu Abwertung führte. Da die Räumlichkeit ohnehin durch den Dekodierprozess gemindert wird, verstärkt sich dieser Effekt bei einer schon von sich aus trockeneren Stereoabmischung zusätzlich.

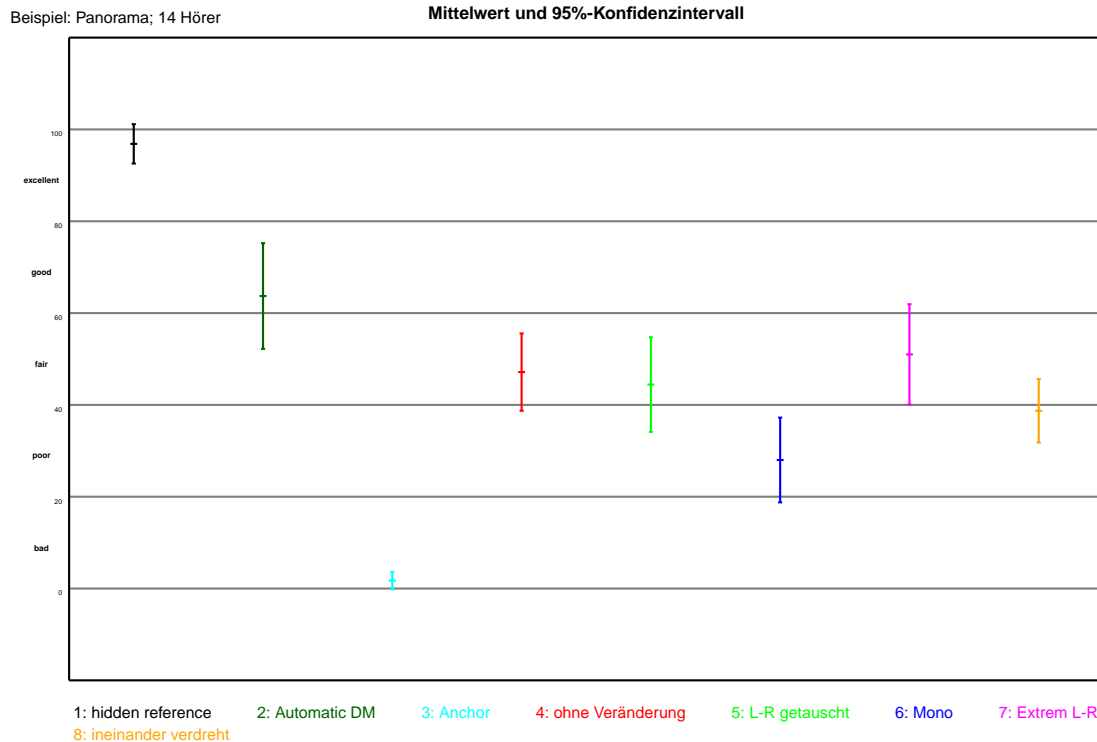


Abbildung 9: Ergebnis des Hörtests zum Parameter „Panorama“

Betrachtet man die dekodierten manuellen Abmischungen für sich, um den Parameter „Panorama“ isoliert zu bewerten, stellt man erstaunlicherweise fest, dass die Abmischung 4 mit der normalen Positionierung und die Abmischung 5, deren Kanäle getauscht waren, nahezu gleich bewertet wurden, d.h. eine komplett gespiegelte Positionierung der Schallquellen in der Stereomischung schien den Dekoder nicht weiter zu beeinträchtigen. Tatsächlich werden die Instrumente völlig korrekt wieder an der ursprünglichen Position abgebildet. Dies lässt sich dadurch erklären, dass der Dekoder die Frequenzanteile gemäß den ICLD wieder an die Position der Originalmischung platziert. Da die Korrelation jedoch nur in der Stereomischung übertragen wird, welche im Fall von Beispiel 5 gespiegelt ist, tritt eine Einengung des Klangbildes auf, da die ursprüngliche Breite der Schallquellen nicht korrekt wiedergegeben werden kann. Dabei geht zusätzlich Raumeindruck verloren, sodass das gesamte Klangbild trockener erscheint. Außerdem kann man eine stärkere Artefaktbildung feststellen, diese schien jedoch nur minimal ins Gewicht zu fallen.

Die gleichen Auswirkungen, allerdings in weitaus stärkerer Form, konnten auch in der dekodierten Surroundfassung der Monomischung, Beispiel 6, festgestellt werden. Da bei diesem Beispiel überhaupt keine ICC übertragen wurde, weder im Stereomix noch in den EBCC-Parametern, verengte sich die Front noch stärker und der Raumeindruck ging nahezu ganz verloren, das Klangbild wurde sehr trocken und extrem phasig. Doch auch in diesem Beispiel waren die Instrumente an den äußersten Positionen, Flöte und Klarinette, korrekt abgebildet. Desweiteren

fiel die Artefaktbildung geringer aus, was vermutlich am fehlendem Hall liegen dürfte.

Ähnlich verhält es sich auch bei Beispiel 8, wo die Instrumente ineinander verdreht positioniert wurden. Mit dem weniger räumlichen Gesamteindruck geht auch hier eine verminderte Artefaktbildung einher. Allerdings ist in diesem Beispiel die Abbildung zwischen vorne und hinten relativ ungenau. So finden sich z.B. von der Flöte nur die hohen Frequenzen, wie sie im Anblasgeräusch enthalten sind, auch wirklich an der Originalposition wieder, ansonsten sind die Instrumente nur schlecht voneinander zu trennen.

Beispiel 7 wiederum zeigt, dass man auch den Parameter „Panorama“ verwenden kann, um Dekodierartefakte zu verkleinern. Durch die extreme Positionierung der Schallquellen nach links oder rechts im Stereobild ergibt sich eine größere Ähnlichkeit zum automatischen Downmix und dadurch eine Positionierung in der dekodierten Surroundfassung, welche der Referenz deutlich näher kommt. Einzige Ausnahme bildet der Bass, welcher in diesem Beispiel auch in den Surroundkanälen erscheint.

6.4 Hall

Die Rolle des Hallanteils in der Stereomischung wurde schon kurz im Unterkapitel „Panorama“ auf Seite 19 angesprochen. Bei diesem Beispiel konnte man schon beobachten, dass eine verminderte Räumlichkeit in der Stereomischung mit einer verminderten Räumlichkeit in der dekodierten Surroundfassung einhergeht. Diese Verminderung hat allerdings keinen nennenswerten Einfluss auf die Funktionsweise des Kodierverfahrens selbst, vielmehr sinkt sogar die Artefaktbildung bei abnehmendem Hallanteil.

Bei dem nun folgendem Beispiel, einer Orgelaufnahme, sollte untersucht werden, welche Auswirkungen qualitativ unterschiedliche Räumlichkeiten in der Stereomischung auf die Surroundmischung haben. Die Aufnahme selbst war eine reine Hauptmikrofonaufnahme, bei der für jeden Surroundkanal jeweils ein Mikrofon verwendet wurde. Bei den Stereomischungen wurden sowohl die Auswirkungen des natürlichen Halls, als auch die Auswirkung künstlichen Halls untersucht. Bei der ersten Mischung wurden die fünf Spuren 1:1 zusammengemischt, bei der zweiten wurden lediglich die vorderen Kanäle benutzt. Bei Beispiel 3 und 4 wurden exemplarisch die künstlichen Hallprogramme „Large Church“ und „Large Hall“ des Hallgeräts „Lexikon 960“ benutzt. Die fünfte Mischung greift nochmal auf die Originalspuren zurück, wobei diesmal die vorderen Kanäle um 3 dB abgesenkt wurden.

An den Ergebnissen des Hörtest, wie in Abbildung 10 dargestellt, kann man sehen, dass die Unterschiede als sehr gering bewertet wurden. In diesem Fall korrelieren die Ergebnisse auch mit den Ergebnissen des Hörtest zu den Stereomischungen (siehe auch Abbildung 5 auf Seite 15), wo der automatische Downmix nahezu gleich wie die manuelle Abmischung bewertet wurde. Sowohl der automatische Downmix als auch die 1:1-Abmischung weisen einen Mittelwert von 56 Punkten auf.

Betrachtet man für beide Beispiele allerdings den in Grafik 11 dargestellten Medianwert sowie die Abmessungen der angrenzenden Quartilen ergibt sich ein leicht anderes Bild. Die manuell erstellte Surroundfassung weist zwar einen gering niedrigeren Zentralwert auf, enthält dafür allerdings mehr Beurteilungen im Bereich „Excellent“. Laut Angaben einiger Teilnehmer des Testfeldes enthielt die automatisch erstellte Surroundfassung zu viel Direktschallenergie auf den hinteren Kanälen, wodurch das Klangbild im Raum sprang. Bei der manuell erstellten Mischung war dies

Beispiel: Hall; 14 Hörer

Mittelwert und 95%-Konfidenzintervall

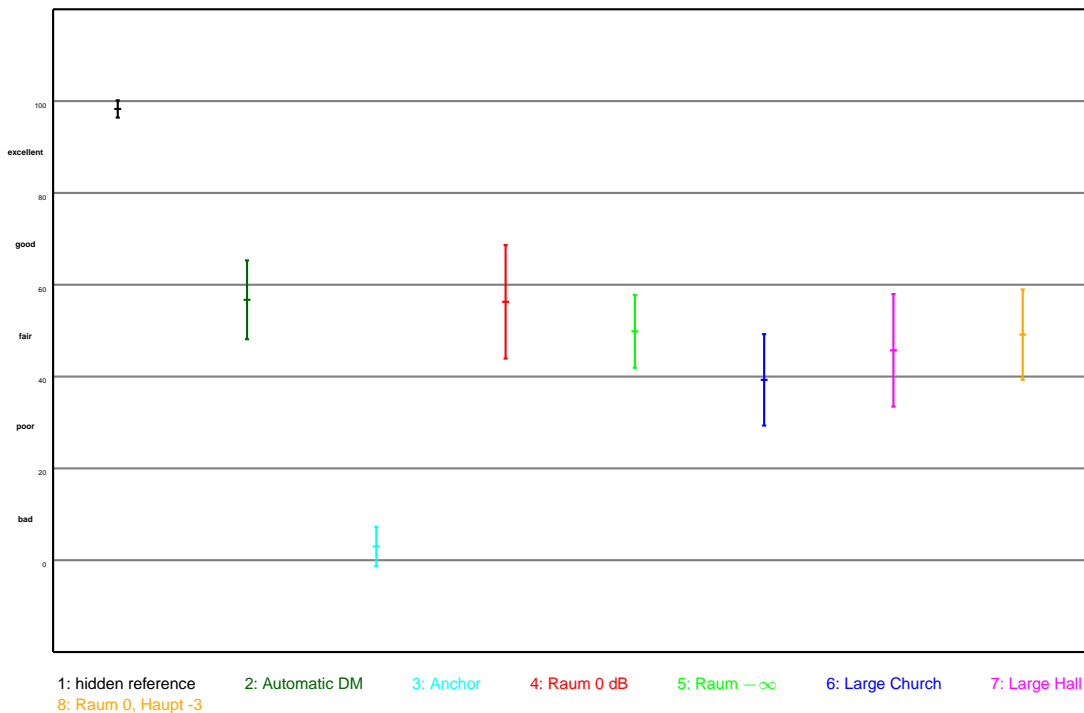


Abbildung 10: Ergebnis des Hörtests zum Parameter „Hall“

nicht der Fall, vermutlich aufgrund der ungleichmäßigen Energieverteilung in der Stereomischung, da der Hallanteil der hinteren Spuren unverändert beigemischt wurde.

Auch die Fassung ohne die Raummikrofone weicht nur marginal um sechs Punkte vom Mittelwert des automatischen Codecs ab. Die Probanden wiesen lediglich auf die Zunahme der Lautstärke in den unteren Frequenzen hin, die dadurch zustande kommt, dass der in der ursprünglichen Surroundmischung „helle“ Raum der hinteren Spuren fehlt. Mit einem Mittelwert von 49 Punkten liegt dieses Beispiel gleichauf mit der Fassung, bei der die vorderen Kanäle um -3 dB abgesenkt wurden. In jenem Beispiel steigt der Bass zwar nicht so stark an, dafür treten jedoch verstärkt Artefakte auf. Durch die Absenkung des Direktschalls der vorderen Kanäle stieg der Hallanteil in der Stereomischung, welcher zu einer größeren und breiteren Räumlichkeit führt. Diese stellt für das BCC-Verfahren insofern eine Herausforderung dar, dass es Schallquellen wahrnehmungstechnisch trennen muss [4]. Dadurch wurde nach Aussage der Probanden das Klangbild „undeutlich“, „schwebend“ und bekam eine „Rauigkeit im Hall“. Somit hoben sich Vor- und Nachteile bei dieser Fassung auf.

Die oben beschriebenen Nachteile traten umso stärker bei der Fassung auf, welche mit dem Kunsthallprogramm „Large Church“ erstellt wurde, wodurch sich das eindeutig schlechte Abschneiden dieses Beispiels erklären lässt. Zum einen wird schon in der Stereomischung der Bass- und Mittenbereich des Halls bis 900 Hz um das 1,6-fache verstärkt, sodass der untere Frequenzbereich beim Dekodierprozess unnatürlich laut wird. Zum anderen weist dieses Hallprogramm mit 3,4 ms eine weitaus größere Nachhallzeit als die Originalmischung auf, wodurch die Räumlichkeit extrem vergrößert wird, was zu den oben genannten Artefakten führt.

Ähnlich verhält es sich mit dem auf dem „Large Hall“-Hallprogramm basierendem Beispiel. Da

bei diesem Hallprogramm sowohl der Bassanstieg (1,2-fach bis 3 kHz) als auch die Nachhallzeit (1,833 ms) nicht ganz so extrem ausfallen, waren die Auswirkungen auf die dekodierte Surroundfassung nicht ganz so gravierend. Allerdings ist bei diesem Beispiel auch im Hallanteil der Stereomischung relativ viel vom Direktsignal enthalten, sodass die räumliche Abbildung unter dem Dekodieren etwas leidet und Direktschall z.T. auch hinten geortet wurde. Dennoch wurde dieses Beispiel nur unwesentlich schlechter als die zweite und fünfte manuelle Abmischung bewertet, woraus geschlossen werden kann, dass MP3-Surround nicht prinzipiell durch Kunsthall beeinträchtigt wird. An dieser Stelle sei explizit erwähnt, dass die obigen Beobachtung keine qualitative Beurteilung der jeweiligen Hallprogramme darstellen sollen.

6.5 Mikrofonierung

Die in den vorherigen Unterkapiteln besprochenen Parameter „Pegel“, „Panorama“ und „Hall“ fließen direkt in den Parameter „Mikrofonierung“ ein, da jede Mikrofonanordnung unterschiedliche Pegel und je nach Mikrofoncharakteristik unterschiedlichen Hallanteil aufzeichnet und somit eine unterschiedliche Abbildung erzeugt. Die Untersuchung zu diesem Parameter richtet sich an Anwender, die für die Surroundfassung und die Stereomischung unterschiedliche Mikrofonierungen verwenden wollen. Um die Anzahl der beeinflussenden Elemente gering zu halten, wurde bei der Erstellung der Hörbeispiele hauptsächlich Augenmerk auf die Mikrofoncharakteristik sowie die Verwendung eines Center-Mikrofons gelegt.

Als Beispiel kam eine Choraufnahme zum Einsatz, bei der die Surroundmikrofonierung mit zwei Kugeln für vorne links und rechts, einer Superniere für den Center und zwei nach hinten gerichteten Nieren für die hinteren Kanäle realisiert wurde. Zusätzlich waren Stützmikrofone im Nahfeld des Chores platziert. Für die Untersuchung wurde an den Kugelmikrofonen noch zwei Nieren des selben Herstellers angebracht, um die verschiedenen Stereomischungen zu erzeugen. Die ersten drei manuellen Stereomischungen entstanden unter Verwendung der Kugeln als Hauptsystem, der Nieren als Hauptsystem und einer Zusammenschaltung der beiden Mikrofonpaare, dem sogenannten Straus-Paket, was rechnerisch eine breite Niere, bzw. Halbkugel ergibt. Für die letzten beiden Beispiele wurde dem Kugel-, bzw. dem Nierenpaar als Hauptsystem zusätzlich noch das Center-Mikrofon beigemischt.

Da schon bei der Untersuchung der oben genannten Einzelparameter erst bei extremen Veränderungen Unterschiede in den dekodierten Surroundfassungen festgestellt werden konnten, verwundert es nicht, dass bei diesem Beispiel die Bewertungen nur geringfügig voneinander abwichen, wie Abbildung 12 zeigt. Die Fassung mit dem Hauptmikrofon bestehend aus den Kugeln und die Fassung mit dem Hauptmikrofon bestehend aus den Nieren wurden mit einem Mittelwert von 55 Punkten exakt gleich bewertet. Das minimal bessere Abschneiden als das des automatisch generierten Mixes (51 Punkte) lässt sich dadurch erklären, dass die Information der Superniere für

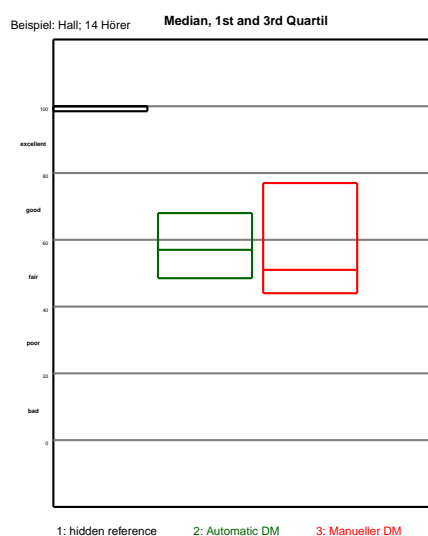


Abbildung 11: Hörtest z. Parameter „Hall“: Median

Beispiel: Mikrofonierung; 16 Hörer

Mittelwert und 95%-Konfidenzintervall

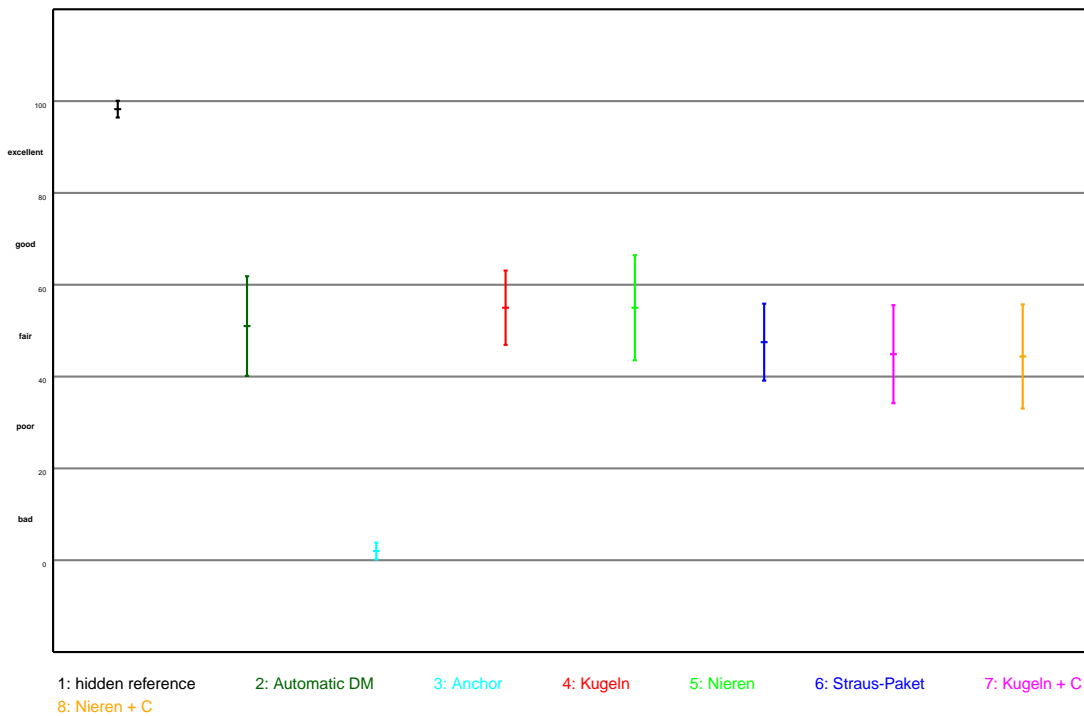


Abbildung 12: Ergebnis des Hörtests zum Parameter „Mikrofonierung“

den Centerkanal nicht in die Mischung einging und dadurch zu einem etwas offenerem Klangbild geführt hat.

Diese sehr direkte Klangkomponente sorgte bei dem vierten und fünften Beispiel zwar für etwas Deutlichkeit in der Abbildungsmitte, verengte dadurch aber noch zusätzlich die Frontalabbildung und verkleinerte den Hallanteil. Ein ähnliches Klangbild ergab sich auch bei dem Beispiel, welches mit dem Straus-Paket erstellt wurde. Der kompakte Klang der Stereomischung führte in der dekodierten Surroundfassung zu einem gleichfalls engen, weniger präsenten und trockenerem Klangbild.

Diese Unterschiede sind jedoch, wie oben bereits erwähnt, nur Nuancen, die generelle Arbeit des Kodierverfahrens wird durch eine unterschiedliche Mikrofonierung nicht beeinträchtigt.

6.6 Delay

Unter dem zu untersuchenden Parameter „Delay“ ist an dieser Stelle die Verzögerung von Stützmikrofonen auf ein Hauptmikrofonsystem gemeint. Dadurch soll verhindert werden, dass von den Stützmikrofonen aufgenommener Direktschall als frühe Reflexion interpretiert wird und somit die Abbildung „verfälscht“. Desweiteren werden durch die Verzögerung Kammfiltereffekte vermindert. Dazu verzögert man die Stützmikrofone genau um den Wert, der sich aus der Entfernung der Mikrofone zueinander in Verhältnis zur Schallgeschwindigkeit ergibt. In einem solchen Fall sind die Mikrofone „auf den Punkt“ verzögert.

Für die Untersuchung wurden von einer Klavieraufnahme neben dem automatischen Downmix noch drei manuelle Downmixe erstellt: einer ohne Verzögerung der Stützmikrofone, einer mit „auf den Punkt“ verzögerten Stützmikrofonen und einer mit zusätzlich um 10 ms¹⁰ verzögerten Stützmikrofonen.

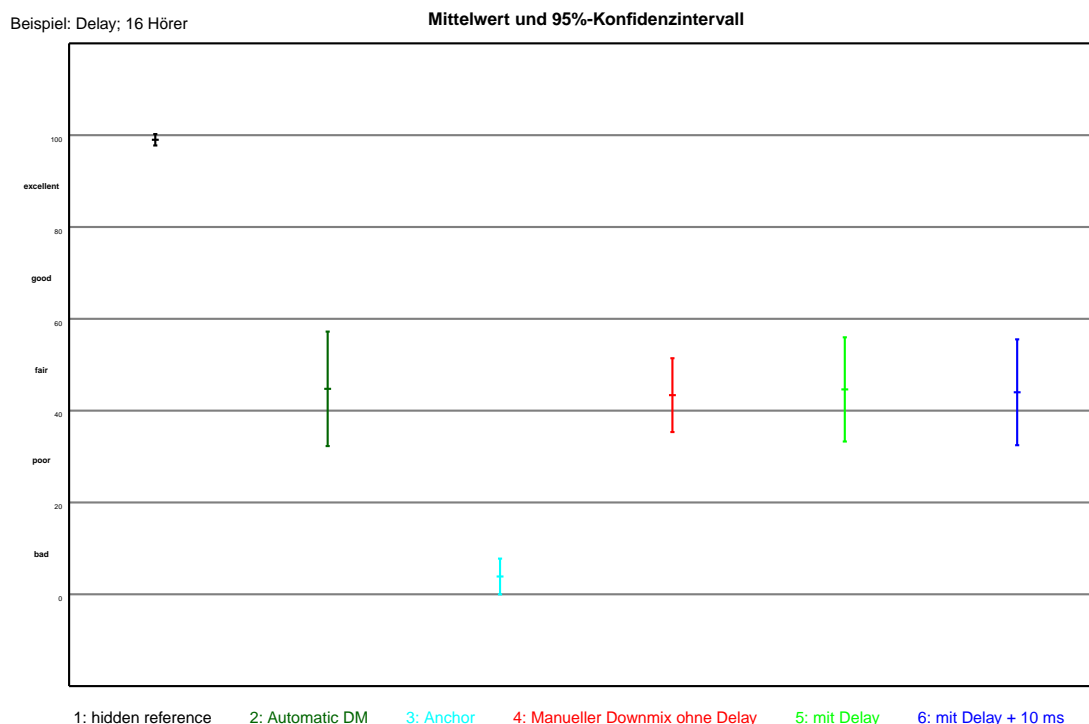


Abbildung 13: Ergebnis des Hörtests zum Parameter „Delay“

Bei keinem Parameter des gesamten Testfeldes kann man mit größerer Sicherheit sagen, dass er keinen Einfluss auf das MP3-Surround-Kodierverfahren zu haben scheint als beim Parameter „Delay“, da alle Beispiele im Durchschnitt gleich bewertet wurden, wie Abbildung 13 zeigt. Die Begründung dafür könnte darin liegen, dass der Einsatz von Delays in der Stereomischung die Frequenzanteile verändert, die für die Frontabbildung verantwortlich sind, weniger aber die Frequenzanteile, welche für die Surroundinformation sorgen. Somit verändert sich der Klang beim Dekodieren gemäß der Klangveränderung in der Stereomischung, wie schon in den vorhergehenden Untersuchungen gezeigt, nicht aber die Räumlichkeit und die Surroundinformation. Das MP3-Surround-Verfahren bleibt also vom Parameter „Delay“ weitestgehend unberührt.

6.7 Dynamics

Bei der Untersuchung des Parameter „Pegel“ auf Seite 17 wurden nur statische Pegeländerungen an den Stereomischungen vorgenommen. Darüber hinaus soll in diesem Unterkapitel der Frage nachgegangen werden, wie MP3-Surround auf dynamische Pegeländerungen reagiert. Dies ist

¹⁰Die zusätzliche Verzögerung um 10ms ist ein in der Tonstudioteknik traditioneller Wert, dessen wissenschaftlicher Hintergrund leider nicht nachweisbar ist. Da dieser Wert von vielen Tonmeister jedoch noch eingesetzt wird, wurde er für diese Untersuchung mit aufgenommen

direkt mit der Frage verknüpft, wie MP3-Surround auf Effekte reagiert, die für solche Anwendungen konzipiert sind. Gemeint sind damit Effekte wie Kompressoren, Limiter, Expander — kurz, die sogenannten Dynamics.

Als Hörbeispiel wurde eine Aufnahme eines Stückes für Schlagwerk ausgewählt, da die percussiven Elemente dieser Aufnahme sowohl für den MP3-Surround-Codec als auch für die Effekte selbst eine Herausforderung darstellten. Eine detaillierte Beschreibung der Effekte sowie deren Einstellung liefert Tabelle 2.

Bsp.	Effekt	Einstellungen
1	ohne Effekt	
2	Limiter	Threshold: -10 dB Delay: 10 ms
3	Kompressor	Threshold: -10 dB Ratio: 1:1,5 Attack: 3 ms / Release: 500 ms Delay: 10 ms
4	2 Kompressoren	Threshold: -20 dB / -10 dB Ratio: 1:2 ; 1:3 Attack: 3 ms / Release: 500 ms Delay: 10 ms
5	Expander	Threshold: - 15 dB Ratio: 2:1

Tabelle 2: Eingesetzte Dynamics

Die Ergebnisse des Hörtest zeigt Abbildung 14. Auffällig ist dabei, dass die Unzulänglichkeiten der Stereomischung, die bei der Beurteilung der Stereofassungen gemäß Abbildung 6 auf Seite 15 zu massiver Abwertung geführt haben, in der Surroundfassung überhaupt nicht ins Gewicht fielen. Die Bewertung der Mischung ohne Einsatz von Dynamics weicht um zu vernachlässigende 0,7 Punkte von der automatisch generierten Fassung ab, die Mischungen mit den Kompressoren wurden sogar besser bewertet.

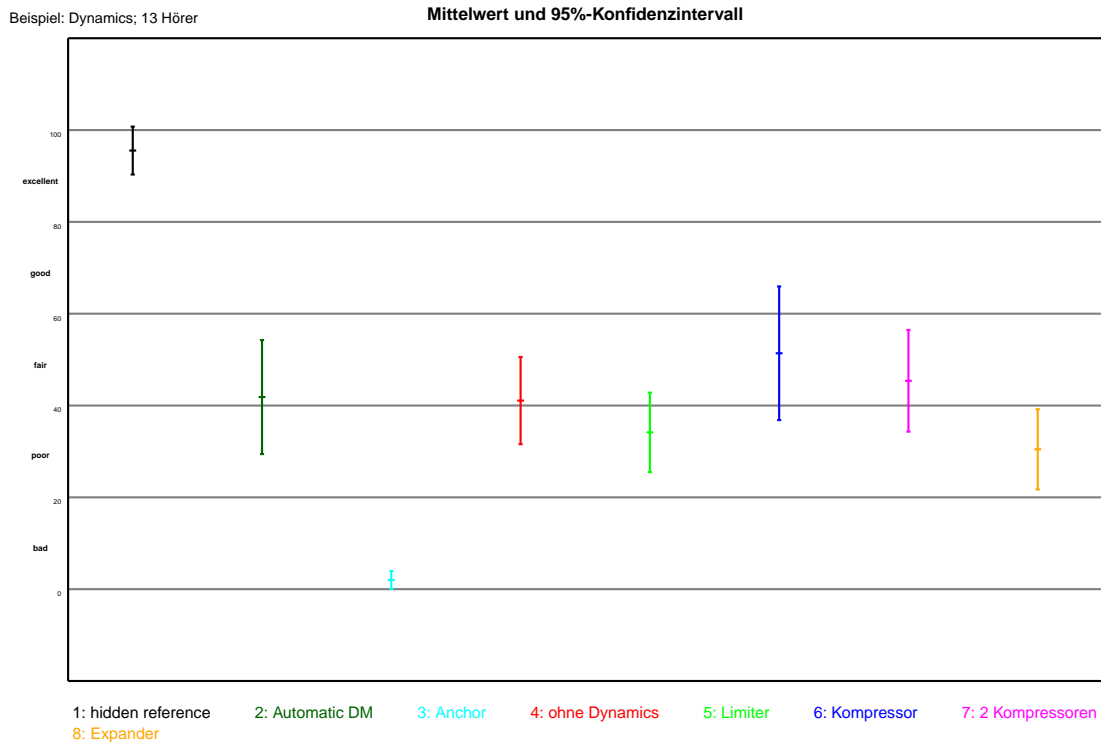


Abbildung 14: Ergebnis des Hörtests zum Parameter „Dynamics“

Der Kompressor verringert den Dynamikumfang einer Aufnahme, sodass bei gleicher Lautstärke

im Vergleich zum automatisch generierten Stereodownmix zwar ein direkteres Klangbild entsteht, effektiv aber auch mehr Energie im Hallanteil der Mischung vorhanden ist. Da die Energieverteilung im Surroundmix relativ zur Energieverteilung in der Stereomischung übertragen wird, bleibt diese Energie beim Dekodieren erhalten und sorgt so für mehr Hall. Damit wird der beim Dekodieren auftretende Verlust von Räumlichkeit abgefangen. Dieser Effekt wird allerdings schnell von klanglichen Verschlechterungen überlagert, die bei stärkerem Einsatz eines Kompressors oder eines Limiters als Extremfall auftreten können. Die Probanden bewerteten sowohl bei Beispiel 5 als auch Beispiel 7 die Klangqualität als „flach“ oder „zu dick“. Beide Kommentare, und dadurch auch die schlechtere Bewertung, lassen sich eher auf den starken Einsatz der Dynamics als auf das MP3-Surround-Verfahren zurückführen.

Etwas anders verhält es sich beim letzten Beispiel, der mit einem Expander bearbeiteten Stereofassung. Der Expander arbeitet in genau entgegengesetzter Richtung wie ein Kompressor. Somit verringert er die Energie im Hallanteil. Dies sorgt schon in der Stereofassung für einen Verlust an Räumlichkeit und wird beim Dekodieren noch zusätzlich verstärkt. Das Gesamtklangbild zerfällt dadurch in Einzelinformationen aus den einzelnen Lautsprechern, denen ein verbindendes Element fehlt.

Prinzipiell jedoch ist der Einsatz von Dynamics für das MP3-Surround-Verfahren relativ unkritisch, kann sogar für den Dekodierprozess förderlich sein.

6.8 Effekte

Der Einsatz von Effekten ist streng genommen kein eigener Parameter, sondern eher eine Kombination der bereits untersuchten. Die Wechselwirkungen zwischen den einzelnen Parametern sind dabei so vielfältig wie die Anzahl der verfügbaren Effekte selbst. Deswegen wurden für die folgende Untersuchung drei Grundtypen von Effekten verwendet: Flanger, Octaver und Phaser. Für das letzte Beispiel wurde eine Kombination von Flanger und Phaser verwendet.

Als Grundlage diente die Surroundmischung eines Rockstückes. Bei den manuell erstellten Stereofassungen wurden die oben genannten Effekte auf eine Rhythmusgitarre und eine Sologitarre angewendet. Die Rhythmusgitarre befand sich in der Surroundmischung im Center, die Sologitarre war zwischen Surround Links und Surround Rechts gepant.

Zuerst sei zu den Ergebnissen des Hörtest, wie Abbildung 15 zeigt, angemerkt, dass die automatisch erzeugte Surroundfassung und die manuell erzeugte, aber ohne Effekte versehene Fassung mit 58,1 und 58,7 Punkten völlig gleich bewertet wurden. Sogar die Standardabweichung differiert um lediglich 1,4 Punkte. Im Vergleich zur Auswertung der Stereoabmischungen auf Seite 15, wo die einzelnen Stereofassungen beim Mittelwert 11 Punkte auseinanderlagen, ist dies besonders zu beachten. Einschränkungen der manuellen Stereomischung scheinen den Dekoder nicht in der Arbeitsweise beeinträchtigt zu haben.

Umso überraschender waren die Ergebnisse des Dekoders bei den bearbeiteten Stereofassungen. Am schlechtesten wurde die Fassung mit dem Flanger bewertet, da dieser Effekt — nachträglich angewendet — zum MP3-Surround-Dekoder inkompatibel ist. Der Flanging-Effekt entsteht durch die Erzeugung künstlicher Nullstellen im Frequenzspektrum mittels Verzögerungsgliedern. Die Nullstellen stehen dabei in einem harmonischen Verhältnis zueinander, ihre Form zeigt einen sinusförmigen Verlauf. Dadurch ergeben sich nur geringe Nichtlinearitäten [10]. Der Flanger moduliert also das Signal über einen relativ großen Frequenzbereich, sodass das entsprechende In-

Beispiel: Effekte; 14 Hörer

Mittelwert und 95%-Konfidenzintervall

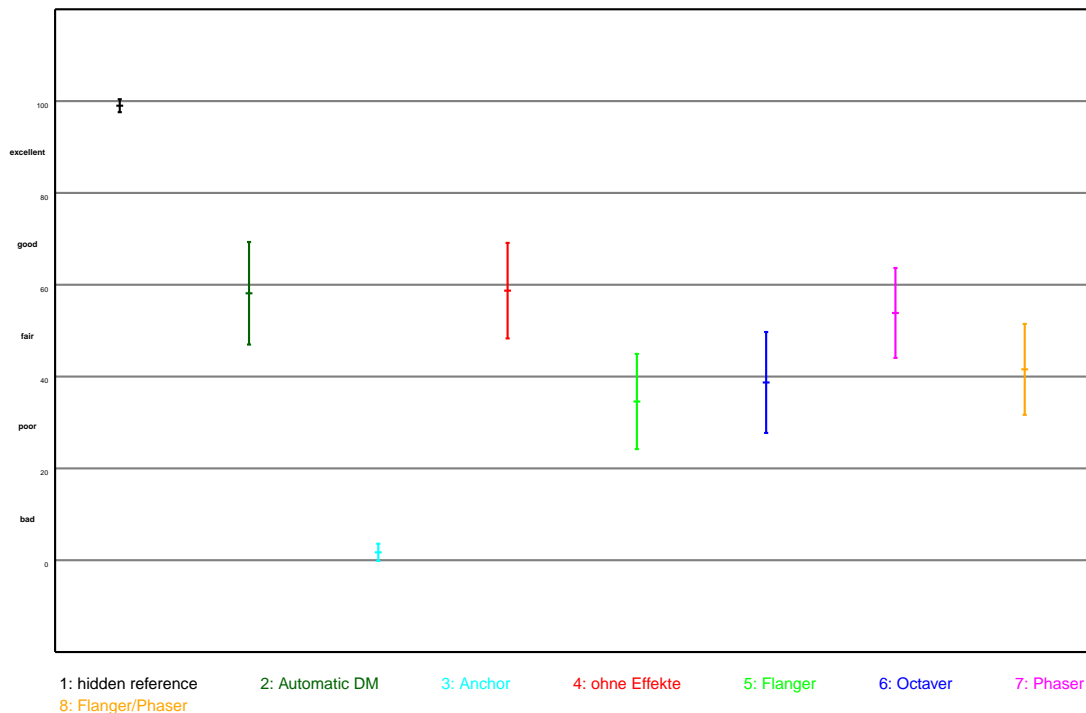


Abbildung 15: Ergebnis des Hörtests zum Parameter „Effekte“

strument plötzlich in Frequenzbändern auftritt, in denen es in der Surroundfassung nicht vorkam. Der MP3-Surround-Dekoder wiederum verteilt die Energie der einzelnen Frequenzbänder gemäß der EBCC-Analyse im Surroundfeld. Dadurch kann es passieren, dass einzelne Frequenzabschnitte des modulierten Signals an unterschiedlichen Positionen platziert werden. Beim vorliegenden Beispiel entstand dadurch eine sich quasi im Surroundfeld „bewegende Schallquelle“, was für einige Probanden „hubschrauberartig“ klang.

Der Einsatz des Phasers schien dagegen weniger Probleme hervorzurufen, was erstaunt, da die beiden Effekte — Flanger und Phaser — vom Klang und der Funktionsweise her relativ verwandt sind. Der Phasing-Effekt entsteht durch Erzeugung künstlicher Nullstellen im Frequenzspektrum durch Phasenverschiebung. Damit stehen die Nullstellen nicht in einer harmonischen Beziehung zueinander und zeigen somit Nichtlinearitäten auf [10]. Der Phaser scheint die Geräuschanteile eines Signals zu betonen, wodurch klanglich der Eindruck entsteht, der Phaser nutze nur die oberen Frequenzen zur Modulation. Dieser Frequenzbereich der Geräuschanteile war bei der vorhandenen Rockmischung fast ausschließlich im Schlagzeug enthalten, welches — genau wie der Phaser selbst — auf die vorderen rechten und linken Kanäle gepant war. Dadurch war der Phaser auch in der Surroundmischung nur in diesen beiden Kanälen zu hören, was allerdings mit 53 Punkten im Vergleich zu den 34 Punkten der Flanger-Fassung als weniger störend bewertet wurde.

Bei der Kombination der beiden Effekte wurde eine hohe Einstellung für die Modulationszeit gewählt, sodass der Effekt in der Stereofassung sehr schnell zwischen links und rechts wechselte. Dieser schnelle Seitenwechsel wurde beim Dekodieren jedoch nicht mittransformiert, sodass der Effekt in der Surroundfassung fast nicht mehr vorhanden war, bzw. lediglich für ein „phasi-

ges“ und „instabiles“ Klangbild der Gitarren verantwortlich war. Dementsprechend wurde dieses Beispiel auch weniger gut bewertet.

Die mit dem Octaver bearbeitete Stereofassung stellt den umgekehrten Fall der beim Parameter „Pegel“ auf Seite 17 dargestellten Szenarien dar. Wurden dort Instrumente entfernt, verändert der Octaver die Stereofassung, als füge man Instrumente hinzu. Dabei gilt letztlich dasselbe Prinzip wie beim Flanger: je nach zugefügtem Frequenzbereich werden diese gemäß der EBCC-Analyse im Surroundfeld wieder platziert. In diesem Beispiel wurden zusätzliche tiefe Frequenzen generiert, die weiterhin in den vorderen Kanälen abgebildet wurden, sodass keine räumlichen Veränderungen auftraten. Lediglich die Solo-Gitarre war plötzlich von zwei Seiten zu hören: von hinten erklang das Original, von vorne die oktavierte Fassung. Die zusätzlichen tiefen Frequenzen beeinträchtigten jedoch schon die Transparenz der Stereofassung und steigerten so die Basszunahme in der dekodierten Fassung auf ein unerträgliches Maß, sodass das gesamte Klangbild noch „dumpfer“ und „undurchsichtiger“ wurde.

7 Zusammenfassung

Bei der vorliegenden Untersuchung hat sich MP3-Surround als ein sehr praxistaugliches und effizientes Kodierverfahren für Mehrkanalaudio bewährt. Die Abwärtskompatibilität zu Stereosystemen ist durch einen guten Downmixalgorithmus gewährleistet, welcher nur minimale Einschränkungen aufweist und statischen Downmixverfahren überlegen ist.

Auch bei Verwendung manueller Abmischungen verhält sich der MP3-Surround-Codec sehr robust. Erst extreme Veränderungen einzelner Parameter, die in der Praxis wohl kaum vorkommen dürften, rufen nennenswerte Verzerrungen hervor, sofern diese nicht sogar vom MP3-Surround-Dekoder kompensiert werden. Das zeigen vor allem die Vergleiche der Hörtests zu den Stereomischungen auf Seite 15 und 15 für die Beispiele Drums und Rock und den entsprechenden Hörtests zu den Surroundmischungen auf den Seite 26 und 28. Gerade bei diesen beiden Beispielen klaffen die Bewertungen der Stereomischungen weit auseinander, die Surroundmischungen jedoch liegen nahezu gleichauf.

Weiterer Untersuchungen bedarf die Qualität der automatisch generierten Stereofassung in Bezug auf Kopfhörerwiedergabe. Dieser Punkt wurde in der vorliegenden Arbeit lediglich durch persönliche Beobachtung angedeutet, jedoch nicht mittels eines Hörtests verifiziert. Außerdem sollten die Auswirkungen des neu implementierten Diffusors untersucht werden.

An dieser Stelle möchte ich mich ganz herzlich bei Herrn Auster und Herrn Soldan von der Firma B&W für die Bereitstellung der Abhöranlage, sowie bei allen Probanden für die Teilnahme an den Hörtests bedanken.

Ist das geschickt, sowas zu schreiben? Diese Liste wird bestimmt noch länger

Literatur

- [1] J. Herre, C. Faller, C. Ertel, J. Hilpert, A. Hölzer, C. Spenger: „MP3 Surround: Efficient and Compatible Coding of Multi-Channel Audio“, 116th AES Convention, Berlin 2004, Preprint 6049

- [2] C. Faller and F. Baumgarte: „Binaural Cue Coding — Part I: Psychoacoustic Fundamentals and Design Principles“, IEEE Transactions on speech and audio processing, Vol. 11, no. 6, November 2003
- [3] T. Painter and A. Spanias: „Perceptual Coding of Digital Audio“, Proceedings of the IEEE, Vol. 88, no. 4, April 2000
- [4] C. Faller and F. Baumgarte: „Binaural Cue Coding — Part II: Schemes and Applications“, IEEE Transactions on speech and audio processing, Vol. 11, no. 6, November 2003
- [5] C. Faller and F. Baumgarte: „Binaural Cue Coding: A novel and efficient representation of spatial audio“, Proc. ICASSP 2002, Orlando, Florida, May 2002
- [6] C. Faller and F. Baumgarte: „Estimation of auditory spatial cues for Binaural Cue Coding“, Proc. ICASSP 2002, Orlando, Florida, May 2002
- [7] C. Faller and F. Baumgarte: „Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compression“. 112th AES Convention, Munich 2002, Preprint 5574
- [8] C. Faller and F. Baumgarte: „Why Binaural Cue Coding is better than Intensity Stereo Coding“. 112th AES Convention, Munich 2002, Preprint 5575
- [9] J. Blauert: *Räumliches Hören*, S. Hirzel Verlag Stuttgart, 1974
- [10] J. Webers: *Das Handbuch der Tonstudioteknik*, 7. Auflage, Franzis' Verlag Poing, 1999
- [11] J. Schwarze: *Grundlagen der Statistik I. Beschreibende Verfahren*, 6. Auflage, Verlag neue Wirtschafts-Briefe, Herne/Berlin
- [12] J. Schwarze: *Grundlagen der Statistik II. Wahrscheinlichkeitsrechnung und induktive Statistik*, 4. Auflage, Verlag neue Wirtschafts-Briefe, Herne/Berlin
- [13] ITU-R Recommendation BS.1534-1: „Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)“, International Telecommunications Union, Geneva, Switzerland, 2001
- [14] ITU-R Recommendation BS.1116: „Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems“, International Telecommunications Union, Geneva, Switzerland, 1994-97

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Detmold, 14. Dezember 2004
