

Diplomarbeit

Downmix-Kompatibilität von Mehrkanalaudio-
Produktionen bei bewegten Signalen unter
Berücksichtigung der Korrelationsbeziehungen

Hochschule für Angewandte Wissenschaften Hamburg
Fakultät Design, Medien und Information
Studiendepartment Technik

vorgelegt von
Volker Wycisk
Gärtnerstr. 89
20253 Hamburg
Matrikel Nr.: 1572 776

Erstgutachterin: Prof. Dr. Eva Wilk
Zweitgutachterin: Katja Zeidler, NDR

Hamburg, den 15. Januar 2006

Versicherung

Ich versichere, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen Quellen, außer den angegebenen, benutzt habe.

Sinngemäße oder wörtliche Übernahmen habe ich angegeben bzw. zitiert.

Hamburg, den 15. Januar 2006

Volker Wycisk

Danksagung:

An dieser Stelle möchte ich all jenen danken, die mich bei der Entstehung dieser Diplomarbeit unterstützt haben:

Simon Jaspersen für die kreative Zusammenarbeit bei der Produktion des Hörspiels.

Herrn Roland Vlaicu von der Firma Dolby für die Kodierung in das Dolby Pro Logic II Format sowie die fachlichen Gespräche und Anregungen.

Meiner Betreuerin Frau Prof. Dr. Wilk für die stets hilfreichen Ratschläge und Gespräche, die mir sehr geholfen haben.

Dem Tonlabor des Fachbereichs Medientechnik und Frau Renate Groth-Lühr für die gute Zusammenarbeit.

Den Schauspielern Felix Maue, Deniz Jaspersen, Helene Matern und Oliver Törner für die professionelle Zusammenarbeit bei der Produktion des Hörspiels.

Sven Kohlwege von der „Mastering & Surround Factory“ im Boogie Park für die Studiozeit und Katja Zeidler für jede Menge Info-Material.

Florian Rathgeber vom ZDF und Roman Velken für die Unterstützung und wertvollen Tipps.

Allen Teilnehmern am Hörvergleich für die Geduld und das „Leihen“ ihrer Ohren.

Besonders möchte ich meinen Eltern danken, die mir das Studium ermöglicht haben und mich durch ihre Korrekturarbeit sehr unterstützt haben.

Meiner Freundin Kirsten Peters danke ich für ihre Unterstützung sowie die Geduld beim Korrekturlesen.

Inhaltsverzeichnis

| | | |
|----------|--|----|
| 1 | Einleitung | 1 |
| 2 | Grundlagen | 2 |
| 2.1 | Problemanalyse..... | 2 |
| 2.2 | Grundlagen Stereophonie | 3 |
| 2.3 | Korrelation und Kohärenz | 7 |
| 2.4 | Addition von Signalen mit einer Phasenverschiebung von 90° | 9 |
| 2.5 | Phantomschallquellen..... | 12 |
| 2.6 | Downmix-Ansätze..... | 13 |
| 3 | Dolby Formate | 14 |
| 3.1 | Formate für den Konsumenten..... | 14 |
| 3.1.1 | Dolby Pro Logic und Dolby Pro Logic II | 14 |
| 3.1.1.1 | Arbeitsweise des Encoders / Decoders..... | 15 |
| 3.1.1.2 | Erstellen einer Lt/Rt-Mischung | 18 |
| 3.1.2 | Dolby Digital..... | 20 |
| 3.1.2.1 | Funktionsweise des Encoders | 21 |
| 3.1.2.2 | Der AC-3 Decoder | 26 |
| 3.1.2.3 | Metadaten | 27 |
| 3.2 | Professionelle Formate - Dolby E..... | 34 |
| 4 | Downmix-Technik | 35 |
| 4.1 | ITU-R 775 Downmix..... | 35 |
| 4.2 | Weitere Möglichkeiten des Downmix | 37 |
| 4.3 | Probleme beim Downmix | 38 |
| 4.3.1 | Technische Fehler..... | 39 |
| 4.3.2 | Abbildung von Räumlichkeit bei der Wiedergabe | 41 |
| 4.3.3 | Richtungsabbildung und Bewegung..... | 44 |
| 4.3.3.1 | Richtungsabbildung und Bewegung mit Hilfe von Pegeldifferenzen | 44 |
| 4.3.3.2 | Richtungsabbildung und Bewegung mit Hilfe von Laufzeitdifferenzen... | 48 |
| 4.3.3.3 | Richtungsabbildung und Bewegung mit Hilfe von Äquivalenzstereophonie..... | 53 |
| 4.3.4 | Aufteilung des Klangfeldes und Auswirkungen auf die Dramaturgie | 55 |

| | |
|---|----|
| 5 Hörvergleich | 57 |
| 5.1 Produktionsbericht: Ein Kurzhörspiel in 3/2-Stereo – „3 x Unschuld“ | 57 |
| 5.1.1 Konzeption und Planung..... | 57 |
| 5.1.2 Vorüberlegungen für die Durchführung der Produktion | 58 |
| 5.1.3 Produktionsablauf..... | 60 |
| 5.2 Durchführung des Hörvergleichs..... | 62 |
| 5.2.1 Audiobeispiele für den Hörvergleich..... | 63 |
| 5.2.2 Versuchsbedingungen | 65 |
| 5.2.3 Versuchsbeschreibung und -durchführung..... | 67 |
| 5.3 Auswertung der Testreihen..... | 69 |
| 5.3.1 Testreihe 1 und 2 | 69 |
| 5.3.2 Testreihe 3 | 74 |
| 5.3.3 Testreihe 4..... | 75 |
| 5.4 Gesamtbetrachtung | 76 |
| 6 Gesamtreflexion und Ausblick | 77 |
| 7 Literatur | 80 |
| 8 Anhang | 86 |

1 Einleitung

In der vorliegenden Diplomarbeit wird die Downmix-Kompatibilität von Mehrkanalaudio-Produktionen bei bewegten Signalen unter Berücksichtigung der Korrelationsbeziehungen untersucht.

Mehrkanalaudio scheint sich langsam für den „Heimgebrauch“ zu etablieren. Dieses liegt zum einen am Siegeszug der DVD – mit dem daraus folgenden Trend zum eigenen Heimkino - und an den immer aufwändiger gestalteten Computerspielen. Zum anderen aber auch daran, dass die Anschaffungskosten für die „Hardware“ stark gesunken sind.

Die Fernseh- und Radioanstalten befinden sich momentan in einer Phase des Ausprobierens und der (Neu-) Orientierung. Nachdem gerade der Sprung ins „digitale Zeitalter“ vollzogen wurde, experimentieren bereits einige Sender mit der neuen/ alten Wiedergabetechnik und suchen nach Lösungen und Einsatzmöglichkeiten für Mehrkanalaudio. Erste Testschleifen des WDR über Astra oder die Sendung „Wetten Dass...?“ im ZDF zeigen, dass Mehrkanalaudio auch im Rundfunk möglich ist.

Diese „Experimente“ sollten möglichst umfangreich ausgeweitet werden, um auch in Zukunft ein konkurrenzfähiges, abwechslungsreiches und anspruchsvolles Programm zu gestalten, mit dem sich evtl. auch neue Hörer für die sogenannten Sparten-Programme gewinnen lassen.

Natürlich müssen auch all diejenigen weiter mit einem adäquaten Klang versorgt werden, denen nicht die Möglichkeiten der Wiedergabe von Mehrkanalaudio zu Verfügung stehen. Und so zeigt sich: Ohne „Stereo“ kein „Surround“.

Mit dieser Arbeit konkretisiere ich die Problematik der Downmix-Kompatibilität von Mehrkanalaudio-Produktionen auf den speziellen Fall der bewegten Signale und gehe dabei folgendermaßen vor:

Die Arbeit gliedert sich in zwei Bereiche. Im ersten Teil erläutere ich die grundlegenden, theoretischen Überlegungen zur Problematik des Downmix. Der zweite Teil der Arbeit ergänzt diese Erläuterungen durch eine praktische Anwendung mittels eines Hörvergleichs.

Zu Beginn der theoretischen Grundlagen gebe ich einen Einstieg in das Thema Downmix und erkläre die für das Verständnis dieser Arbeit grundlegenden Begrifflichkeiten und Definitionen. Dieses Kapitel beschäftigt sich mit den Grundlagen der Korrelation, Phantomschallquellen und dem ITU-R775-Standard zur Wiedergabe von 3/2-Stereosignalen.

Anschließend beschreibe ich im dritten Kapitel, anhand eines Einblicks in die unterschiedlichen Audioformate der Firma Dolby, zwei Varianten des automatischen Downmix und die Anwendung innerhalb des Dolby Digital Standards.

Eine vertiefte Analyse des Themas folgt in Kapitel vier. Neben dem Downmix nach ITU-Standard ITU-R 775 als eine weitere Variation des automatischen Downmix werden technische Fehler sowie die Auswirkungen des Downmix auf das Programm diskutiert. Neben der Abbildung von Räumlichkeit geht es in diesem Kapitel außerdem um die Richtungsabbildung und Bewegung von Phantomschallquellen im 3/2-Stereobild, sowie die Veränderungen, die durch die Reduzierung der Audiokanäle hervorgerufen werden.

Mit Hilfe eines Hörvergleichs werden die theoretischen Überlegungen aus dem Kapitel vier in der Praxis untersucht. Kapitel fünf beschreibt die Durchführung und die Ergebnisse des Hörvergleichs, der die Auswirkungen des Downmix von bewegten Phantomschallquellen anhand von unterschiedlichen Audiobeispielen darlegen soll.

Abschließen stelle ich die Ergebnisse in einen Gesamtzusammenhang und gebe einen Ausblick hinsichtlich der zukünftigen Anwendung.

2 Grundlagen

In diesem Kapitel sollen die für das Verständnis der vorliegenden Arbeit grundlegenden Begriffe und Standards erläutert werden.

Die Begriffe und Zusammenhänge der Audiosignalverarbeitung werden dabei als bekannt vorausgesetzt und werden daher nicht erläutert. Zur Einführung und Vertiefung dieser Sachgebiete wird auf die Arbeiten von Dickreiter, Katz, Pohlman und Watkinson verwiesen. Als grundlegende Lektüre der Mehrkanaltonotechnik empfehle ich Birkner.

2.1 Problemanalyse

Ein *Downmix* ist die Erstellung eines n-kanaligen Audiosignals aus einem Audiosignal mit mehr als n Kanälen. Dieser Vorgang kann „per Hand“, d.h. durch das Erstellen einer neuen Mischung oder mit technischen Hilfsmitteln automatisch erfolgen. Der Vorteil des automatischen Downmix ist, dass dieser im Gegensatz zur erneuten Gestaltung der Mischung wesentlich geringere Kosten verursacht, und damit die Produktion von Mehrkanalaudio nicht zwingend auch mit höheren Gesamtkosten verbunden ist.

Da das Thema des automatischen Downmix in der Vergangenheit schon unter unterschiedlichen Gesichtspunkten betrachtet wurde, soll in diese Arbeit untersucht

werden, wie sich ein automatischer Downmix auf Audiosignale auswirkt, die sich von dem bisher diskutierten Programmmaterial, wie Konzerten oder Showformaten, unterscheiden. Als Ausgangspunkt dient hierzu die Beschäftigung mit dem Thema der Produktion von Hörspielen in Mehrkanalaudio und die damit verbundenen Ideen und Wünsche befreundeter Hörspiel-Regisseure.

Für Künstler aus diesem Gebiet bedeutet die Mehrkanaltechnik oft die Möglichkeit einer realistischeren Darstellung ihrer Klangbilder und Vorstellungen. Einige der ersten Ideen sind dabei stets, neben einer den Hörer einhüllenden Atmosphäre, vorbeifahrende Autos, Flugzeuge, die über die Köpfe der Hörer „donnern“, oder andere Bewegungen im Klangbild. Diese Effekte und Bewegungen sind dabei nicht einfach zu realisieren, da oft die Möglichkeiten der Klangabbildung der gängigen Mehrkanalformate überschätzt werden.

Der Wunsch von Realität und Bewegung scheint dabei vor allem aus der Erfahrung mit Kinofilmen zu stammen, bei denen Mehrkanalaudio schon seit langem kein Novum mehr darstellt. Doch wer genau hinhört, wird feststellen, dass die Raumschiffe und andere Fahrzeuge auch hier nur unter besonderen Bedingungen so zu lokalisieren sind, wie dieses in der eigenen Erinnerung der Fall zu sein scheint.

Es zeigt sich außerdem in der Produktion solcher Spezialeffekte, dass ein genaues Abbild der Wirklichkeit in der Regel zu enttäuschenden Ergebnissen führt. Für die Gestaltung von besonderen Klangeindrücken und Spezialeffekten ist auch bei der Mehrkanaltechnik ein genaues Hinhören und Herausarbeiten der gewünschten Eigenschaften des Audiomaterials nötig.

Unter welchen Voraussetzungen die Erstellung dieser „Bewegungen“ möglich ist, und welche Auswirkungen der Downmix auf diese Audiosignale hat, wird in dieser Arbeit untersucht.

2.2 Grundlagen Stereophonie

Mit dem Begriff Mehrkanal-Stereophonie sollen in dieser Arbeit alle Wiedergabeformate bezeichnet werden, die über mehr als zwei Kanäle verfügen. Der Begriff Surround hingegen beschreibt lediglich die akustische Eigenschaft eines Wiedergabesystems, das den Hörer in das Klanggeschehen „einhüllt“ oder ihn mit Klang umgibt.

Zur Bezeichnung von Mehrkanal-Systemen haben sich zwei Varianten durchgesetzt. Die erste Variante bezeichnet zunächst die Anzahl derjenigen Audiokanäle, die den vollen Frequenzumfang aufweisen und dann getrennt durch einen Punkt die Anzahl der

Audiokanäle mit begrenztem Frequenzumfang, z.B.: 5.1. Die zweite Variante trennt die Anzahl der vor dem Hörer platzierten Audiokanäle durch einen Schrägstrich von der Anzahl der hinter oder seitlich zum Hörer platzierten Audiokanäle, z.B. 3/1. Im Frequenzumfang begrenzte Audiokanäle werden durch eine Zahl hinter einem weiteren Schrägstrich kenntlich gemacht, z.B.: 3/2/1. Diese Art der Schreibweise werde ich im folgenden verwenden.

| System Signale / | Kanäle | Variante 1 | Variante 2 |
|--|---------------------------------------|-------------------|-------------------|
| Mono | M | 1.0 | 1/0 |
| Zweikanal-Stereo | L,R | 2.0 | 2/0 |
| 2 Frontkanäle 1 Surroundkanal | L, R, MS | 3.0 | 2/1 |
| 2 Frontkanäle 2 Surroundkanäle | L, R, LS, RS | 4.0 | 2/2 |
| Dreikanal-Stereo | L, C, R | 3.0 | 3/0 |
| 3 Frontkanäle 1 Surroundkanal | L, C, R, MS | 4.0 | 3/1 |
| 3 Frontkanäle 2 Surroundkanäle | L, C, R, LS, RS | 5.0 | 3/2 |
| 3 Frontkanäle 2 Surroundkanäle 1 Effektkanal | L, C, R, LS, RS, LFE | 5.1 | 3/2/1 |
| 5 Frontkanäle 2 Surroundkanäle | L, LC, C, RC, R, LS, RS | 7.0 | 5/2 |
| 5 Frontkanäle 2 Surroundkanäle 1 Effektkanal | L, LC, C, RC, R, LS, RS, LFE | 7.1 | 5/2/1 |
| 5 Frontkanäle 4 Surroundkanäle | L, LC, C, RC, R, LS1,RS1, LS2, RS2 | 9.0 | 5/4 |

Tabelle 2-1 : Mehrkanalformate und deren Bezeichnung¹

¹ L = linker Audiokanal; R = rechter Audiokanal; C = Center; MS = Mono-Surroundkanal ; LS = linker Surroundkanal; RS = rechter Surroundkanal; etc.

Um eine Wiedergabe auf unterschiedlichen Systemen zu gewährleisten, wurde von der International Telecommunication Union (ITU) eine Hierarchie von Wiedergabesystemen definiert, die eine auf- und abwärtskompatible Umwandlung der verschiedenen Formate ermöglicht. Die Tabelle 2-1 gibt einen Überblick über mögliche Systeme und deren Bezeichnung in den beiden Varianten (vgl. Stevaux, S.7).

Der für ein Wiedergabesystem von 3/2-Stereo international maßgebliche Standard wird von der ITU unter der Bezeichnung „ITU-R BS. 775-1: Multichannel Stereophonic Sound System with and without accompanying Picture“ herausgegeben. Die Grundzüge dieser Empfehlung sollen hier kurz angesprochen werden.

Es wird davon ausgegangen, dass für die Mehrzahl der Wiedergabefälle das bevorzugte Hör- und Sehfeld im vorderen Bereich liegt. Das Mehrkanalsystem besteht aus drei Frontkanälen Links, Center und Rechts, die um zwei rückwärtige, seitliche Surroundkanäle ergänzt werden. Die Kanäle links und rechts werden in einem Winkel von $\pm 30^\circ$ zur Blickrichtung eines Hörers aufgestellt. Es entsteht so ein gleichseitiges Dreieck zwischen den Lautsprechern und dem Hörer. Diese Aufstellung entspricht auch der 2/0-Stereo-Standard-Aufstellung, weshalb diese hier nicht weiter erläutert wird.

Der Centerkanal sorgt dabei für eine Verbesserung der Mittenstabilität der Audiowiedergabe und führt somit zu einer Vergrößerung der Hörzone gegenüber der 2/0-Stereo-Wiedergabe. Zum anderen ist so eine Reproduktion des Dialogkanals bei filmbegleitendem Audiomaterial gewährleistet. Die Surroundkanäle sind ursprünglich für die Wiedergabe von räumlichen Audioinformationen konzipiert worden.

Alle Lautsprecher sollen in idealer Weise auf einer Kreislinie angeordnet werden, in deren Mittelpunkt sich der Platz des besten Hörens – der „Sweet-Spot“- befindet. Ist eine solche Aufstellung nicht möglich, sind die betroffenen Lautsprecher entsprechend zu verzögern.

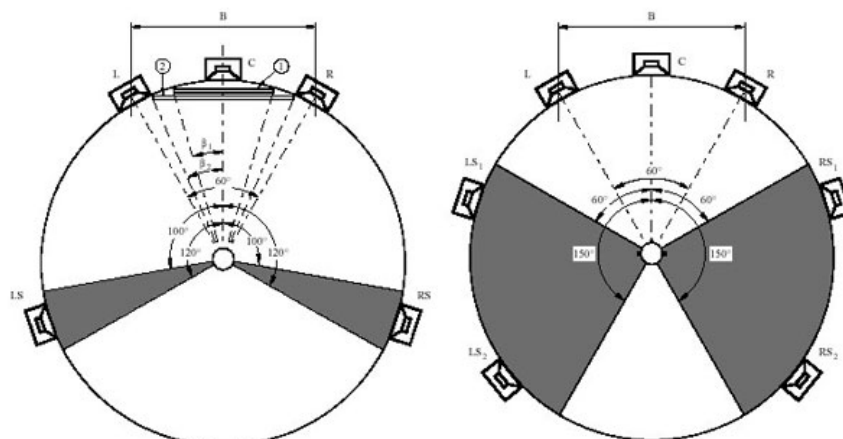


Abbildung 2-1: ITU-R775 Standard-Aufstellung der Lautsprecher

Für die Anordnung der Surround-Lautsprecher wird laut Standard eine Aufstellung seitlich hinter dem Hörer empfohlen, so dass sie mit einer Geraden zum Abhörpunkt einen Winkel von 110° zur Geraden des Center-Lautsprechers bilden.

Eine Option des Standards empfiehlt, zur Verbesserung der Einhüllung zwei zusätzliche Lautsprecher im hinteren Bereich anzuordnen, wo dies möglich ist. Die Signale dafür können aus den beiden Surroundsignalen über einen Prozessor abgeleitet werden. Andererseits kann man mit diesem Prozessor auch ein Signal für einen hinteren Mittenlautsprecher ableiten und diesen bei 180° anordnen (vgl. Steinke, S.9).

Ein Trend zur leichten Abwandlung der Surround-Lautsprecher-Aufstellung lässt sich in Produktions-Studios beobachten, die nicht vornehmlich klassische Musik produzieren. Hier werden die Surround-Lautsprecher bevorzugt in einem Winkel von 125° aufgestellt, um einer Kinoanordnung näher zu kommen (vgl. Mitchel; [P&E], S. 1-3).

Eine weitere Option ist der aus dem Filmbereich stammende LFE-Kanal² für den Tieftonbereich von 20 bis maximal 120 Hz. Beim Film wurde er von der SMPTE³ 1989 als verbindlich eingeführt und kann als Ergänzung für tieffrequente Effekte genutzt werden. Der LFE-Kanal wird oft als Subwoofer⁴ zur Wiedergabe der niedrigen Frequenzen missverstanden. Im Kinobereich wird das Signal des LFE-Kanal immer über einen separaten Subwoofer-Lautsprecher wiedergegeben; hier sind also LFE und Subwoofer direkt im Zusammenhang zu sehen. Der Einsatz eines LFE-Kanals ist jedoch nach ITU für Studio und Heimkinoformate ausdrücklich optional, so dass dieser keine Signale enthalten soll, die für die Darstellung des Gesamtklangs von großer Bedeutung sind. Die Annahme, dass der LFE allein für die Wiedergabe von tieffrequenten Anteilen zuständig ist, kommt vermutlich daher, dass viele der Wiedergabelösungen für Heimanwender zur besseren Wiedergabequalität, Frequenzen unter maximal 120 Hz nur über den LFE-Kanal wiedergeben. Diese Option nennt sich *Bassmanagement*. Hierzu werden die tieffrequenten Anteile der 3/2-Stereo-Mischung mit Hilfe von Filtern getrennt und dem LFE-Kanal hinzuaddiert.

Für weitere Informationen und Empfehlungen für die Praxis, verweise ich z.B. auf die Homepage des Verbands Deutscher Tonmeister unter der Rubrik „Forum“ → „Surround“.

² LFE = Low Frequency Enhancement bzw. Extension oder Effect.

³ Society of Motion Picture and Television Engineers

⁴ Als „Subwoofer“ soll hier ein Lautsprecher-System zur Wiedergabe von ausschließlich tieffrequenten Signalanteilen bezeichnet werden. Diese Einrichtung ist lediglich auf der Wiedergabeseite zu finden und ist damit unabhängig vom Tonformat.

2.3 Korrelation und Kohärenz

Als Grundlage für die Entstehung von (bewegten) Phantomschallquellen werden in diesem Abschnitt die Begriffe Korrelation und Kohärenz erläutert.

Die *Korrelation* von zwei Signalen gibt den Grad ihrer Verwandtschaft oder Ähnlichkeit an. Der *Korrelationsfaktor* ist demnach ein Maß für die Ähnlichkeit zweier Signale oder Funktionsverläufe.

Man unterscheidet bei der Analyse von Signalen zwischen Energie- und Leistungssignalen. Man spricht von Energie- und Leistungssignalen, wenn E bzw. P als nichtnegative endliche Werte existieren. Energiesignale sind alle zeit- und amplitudenbegrenzte Signale, wie z.B. ein Rechteckimpuls. Leistungssignale sind periodische Signale, wie die Sinusfunktion. (vgl. Werner, S. 17-18)

Bei der Überlagerung von zwei Energiesignalen $x(t)$ und $y(t)$ berechnet sich die Energie des resultierenden Signals $s(t) = x(t) + y(t)$ wie folgt:

$$E_s = \int_{-\infty}^{\infty} s^2(t) dt = \int_{-\infty}^{\infty} [x(t) + y(t)]^2 dt = \underbrace{\int_{-\infty}^{\infty} x^2(t) dt}_{E_x} + \underbrace{\int_{-\infty}^{\infty} y^2(t) dt}_{E_y} + 2 \underbrace{\int_{-\infty}^{\infty} x(t) \cdot y(t) dt}_{E_{xy}}$$

E_x ist die Energie von $x(t)$. E_y ist die Energie von $y(t)$. E_{xy} wird mit Kreuzenergie bezeichnet und über die folgende Gleichung mit den Signalenergien in Beziehung gesetzt:

$$E_{xy} = \int_{-\infty}^{\infty} x(t) \cdot y(t) dt = k \sqrt{E_x \cdot E_y} .$$

Mit k wird der Korrelationsfaktor bezeichnet. Für ihn gilt:

$$-1 \leq k \leq +1$$

In der Audiotechnik haben wir es jedoch vor allem mit Leistungssignalen zu tun. Die resultierende mittlere Leistung P_{ges} zweier Signale $x(t)$ und $y(t)$ berechnet sich als:

$$P_{ges} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [x(t) + y(t)]^2 dt = \underbrace{\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x^2(t) dt}_{P_x \geq 0} + \underbrace{\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T y^2(t) dt}_{P_y \geq 0} + \underbrace{\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) \cdot y(t) dt}_{2P_{xy}}$$

P_x ist die mittlere Leistung von $x(t)$ und P_y ist die mittlere Leistung von $y(t)$. Die Größe P_{xy} heißt Kreuzleistung.

Für sie gilt:

$$P_{xy} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) \cdot y(t) dt = \bar{k} \sqrt{P_x \cdot P_y} .$$

Der Faktor \bar{k} wird als Kreuzleistungsfaktor bezeichnet.

Für den Kreuzleistungsfaktor gilt in Analogie zum Korrelationsfaktor ebenfalls :

$$-1 \leq \bar{k} \leq +1$$

Wenn $\bar{k} = 0$ ist, dann nennt man die Leistungssignale $x(t)$ und $y(t)$ orthogonal. Die Kreuzleistung ist dann $P_{xy} = 0$.

Es lässt sich zeigen, dass bei Überlagerung zweier gleichfrequenter, aber um den Winkel ρ verschobener Sinusschwingungen der Korrelationsfaktor $k = \cos \rho$ ist.

Der Korrelationsfaktor hängt damit von der zeitlichen Verschiebung der beiden Signale ab. Zur qualitativen Erfassung der Abhängigkeit des Korrelationsfaktors von der zeitlichen Verschiebung hat man die Korrelationsfunktionen eingeführt.

Die *Kreuzkorrelationsfunktion* gibt dabei die korrespondierenden Eigenschaften von zwei Signalen wieder und ist für zeitkontinuierliche Leistungssignale definiert durch:

$$k_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} x(t) \cdot y(t + \tau) dt \quad \text{mit } T = \text{Beobachtungszeit und } \tau = \text{Verzögerungszeit}$$

Wird statt zweier unterschiedlicher Signale ein Signal mit „sich selbst“ bei unterschiedlichen Verzögerungszeiten verglichen, wird die sich hierbei ergebene Korrelationsfunktion als Autokorrelationsfunktion bezeichnet. Die grafische Darstellung der Autokorrelationsfunktion ergibt an der Stelle $\tau = 0$ unabhängig vom zeitlichen Verlauf des zu analysierenden Signal stets ein positives Maximum.

Zwei Signale werden mit *kovariant* bezeichnet, wenn gilt: $0 < k(\tau)$. Die Signale sind gleichläufig und bei $k(\tau) = 1$ maximal korreliert.

Die Signale sind *kontravariant*, wenn gilt: $k(\tau) < 0$. Die Signale sind gegenläufig und antikorreliert bei $k(\tau) = -1$.

Bei einem $k(\tau) = 0$ sind die Signale maximal unähnlich, stehen also *orthogonal* zueinander. (vgl. Ruprecht, S.67-75; S.223-232; Veit, S.116-119)

Bei einer Mikrophonaufnahme ist die Korrelation von Signalen z.B. abhängig von der Mikrofonbasis, der Beschaffenheit des Raumes – damit von der Art des Diffusschalls – sowie dem Zusammenspiel dieser Faktoren. Der in der Audiotechnik verwendete Korrelationsgradmesser gilt daher nur im Zusammenhang mit Intensitätsstereophonie oder bei Produktionen, die mit Hilfe von Einzelmikrofonverfahren und Pan-Pot erstellt wurden.

Der Korrelationsgradmesser zeigt nicht den momentanen Augenblickswert der Phasendifferenz zwischen dem linken und rechten Signal einer 2/0-Stereoaufnahme an,

sondern einen Wert, der über einen Zeitraum von ca. 0,5 Sekunden integriert ist. Die Skala des Korrelationsgradmessers zeigt nur für Sinus-Messtöne die Kosinuswerte der Phasendifferenz zwischen den Audiokanälen an: $k = \cos \varphi$ ⁵ (vgl. Sengpiel, Gedanken zum Korrelationsgradmesser). Der Korrelationsgradmesser dient somit zur Kontrolle der Monokompatibilität bei Intensitätsstereophonie. Ein Hauptgrund für eine Kontrolle ist die Tatsache, dass Phasenverschiebungen zwischen dem linken und rechten Audiokanal bei der elektrischen Addition zu einem Monokanal zu Auslöschungen führen können.

Als *Kohärenzgrad* K wird der Wert des absoluten Maximums des Betrages der normierten Kreuzkorrelationsfunktion zweier Signale $x(t)$ und $y(t)$ bezeichnet:

$$K = \max_{\tau} |k_{xy}(\tau)| \quad (\text{vgl. Blauert, 1974, S.190})$$

Der Kohärenzgrad hängt also direkt mit der Korrelation zusammen. Kohärent sind dabei Signale deren Kohärenzgrad gleich 1 ist:

$$\max_{\tau} |k_{xy}(\tau)| = 1$$

Entsprechend dieser Festlegung sind Signale also dann kohärent, wenn sie identisch sind oder einen oder mehrere der folgenden Unterschiede aufweisen:

1. unterschiedliche Amplitude bei gleicher Kurvenform, d.h. eine frequenzunabhängige Pegeldifferenz
2. gegenseitige verzerrungsfreie Verzögerung, d.h. Unterscheidung durch eine frequenzunabhängige Phasenlaufzeit
3. gegenseitige Inversion, d.h. gegenseitige Unabhängige Phasendifferenz von 180°

Signale, die sich auf eine andere der genannten Arten unterscheiden, werden als teilkohärent oder inkohärent bezeichnet (vgl. Blauert, 1974, S. 162; Gernemann).

2.4 Addition von Signalen mit einer Phasenverschiebung von 90°

Einige der Verfahren, mit deren Hilfe ein automatischer Downmix erstellt werden kann, verschieben bei bestimmten Audiosignalen die Phasenlage um 90°, bevor die einzelnen Audiokanäle zu einem Zweikanal-Audiosignal zusammengemischt werden. Die Auswirkungen der Addition von zwei Signalen, die in der Phase um 90° verschoben sind,

⁵ Also: $\cos 0^\circ = +1$; $\cos 30^\circ = +0,5$; $\cos 90^\circ = 0$; $\cos 120^\circ = -0,5$ und $\cos 180^\circ = -1$.

sollen daher an dieser Stelle zum besseren Verständnis der nachfolgenden Überlegungen erläutert werden.

Bei der Addition von zwei Audiokanälen kann eine durch die elektrische Addition entstehende Anhebung der Signale in der Stereomitte von 3 dB mittels einer frequenzunabhängigen Phasenverschiebung von 90° eines Kanals vermieden werden.

Strahlen z.B. bei 2/0-Stereo-Standard-Aufstellung der linke und rechte Lautsprecher ein Signal von 0 dB ab, ergibt sich dabei für das Ohr ein Leistungssummenpegel von:

$$10dB \cdot \log(L^2 + R^2) = 10dB \cdot \log(2) = +3dB \quad \text{mit } L = R = 1 \text{ V} \rightarrow 0 \text{ dB}$$

Nach der elektrischen Addition ergibt sich jedoch ein Pegel von:

$$20dB \cdot \log(L + R) = 20dB \cdot \log(4) = +6dB$$

Am Beispiel des Pegels für die Überlagerung von zwei Sinussignalen gleicher Frequenz sei der Effekt einer Phasenverschiebung von 90° demonstriert.

Die resultierende mittlere Leistung zweier Sinussignale

$$u_1(t) = \hat{u} \cdot \sin(\varpi \cdot t + \varphi_1) \quad \text{und} \quad u_2(t) = \hat{u} \cdot \sin(\varpi \cdot t + \varphi_2),$$

ergibt sich aus den vorangegangenen Überlegungen zu:

$$P = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [u_1(t) + u_2(t)]^2 dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u_1^2(t) dt + \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u_2^2(t) dt + \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u_1(t) \cdot u_2(t) dt$$

Mit einem $\Delta\rho = 0^\circ$ folgt daraus:

$$u_1(t) = u_2(t) = \hat{u} \cdot \sin(\varpi \cdot t) \Rightarrow P_{ges} = 4 \cdot \lim_{T \rightarrow \infty} \frac{\hat{u}^2}{2T} \int_{-T}^T u_1^2(t) dt = \underline{\underline{2\hat{u}^2}}$$

Mit $P_1 = \lim_{T \rightarrow \infty} \frac{\hat{u}^2}{2T} \int_{-T}^T u_1^2(t) dt = \frac{\hat{u}^2}{2}$ ergibt sich daher für die beiden kovarianten Signale:

$$10dB \cdot \log \frac{P_{ges}}{P_1} = \underline{\underline{+6dB}}.$$

Sind die beiden Signale orthogonal zueinander ist $\Delta\rho = 90^\circ$, so gilt:

$$u_1(t) = \hat{u} \cdot \sin(\varpi \cdot t) \quad \text{und} \quad u_2(t) = \hat{u} \cdot \cos(\varpi \cdot t)$$

Daraus folgt:

$$P_{ges} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [u_1(t) + u_2(t)]^2 dt$$

$$P_{ges} = \lim_{T \rightarrow \infty} \frac{\hat{u}^2}{2T} \int_{-T}^T \sin^2(\omega t) dt + \lim_{T \rightarrow \infty} \frac{\hat{u}^2}{2T} \int_{-T}^T \cos^2(\omega t) dt + \lim_{T \rightarrow \infty} \frac{\hat{u}^2}{2T} \int_{-T}^T \sin(\omega t) \cdot \cos(\omega t) dt$$

$$P_{ges} = \frac{\hat{u}^2}{2} + \frac{\hat{u}^2}{2} = \underline{\underline{\hat{u}^2}}$$

Daraus ergibt sich:

$$10dB \cdot \log \frac{P_{ges}}{P_1} = \underline{\underline{+3dB}}$$

Das Ergebnis ist eine um 3 dB geringere Leistungssumme bei orthogonalen Signalen. Und somit nähert sich die Addition der natürlichen Wahrnehmung besser an. Die Signale sind also durch die Phasenverschiebung von 90° dekorreliert und stehen nun orthogonal zueinander.

Da die Phasenverschiebung von 90° im Zusammenhang mit dem Downmix später eine wichtige Rolle spielt, wird das gleiche Beispiel noch einmal auf andere Weise erläutert: Durch die Addition von zwei kohärenten Sinussignalen ergänzen sich bei einem $\Delta\rho = 0^\circ$ immer die Maxima der Kurven. Bei einem $\Delta\rho = 90^\circ$ hingegen bleibt die Gesamtleistung gleich.

Durch die Phasenverschiebung von 90° kann außerdem die Auslöschung von Signalen vermieden werden, die

eine Phasendifferenz von 180° aufweisen. Leider kann es aber auch zu neuen Auslöschungen kommen, wenn die Signale erst durch die Filterung eine Phasendifferenz von 180° erhalten. Bei Signalen ohne festen Phasenbezug hat die Phasenverschiebung von 90° hingegen keinen Einfluss auf die Summenbildung.

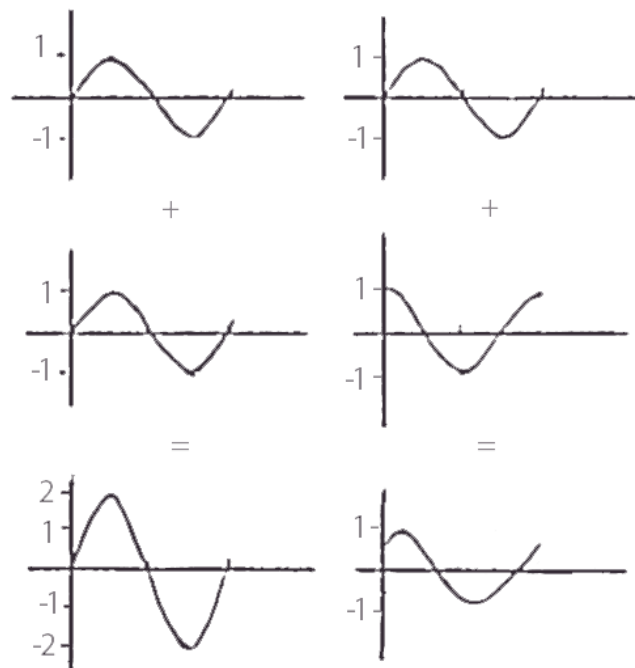


Abbildung 2-2 : Addition von Sinussignalen

2.5 Phantomschallquellen

Die Voraussetzung für das Entstehen von bewegten Audiosignalen bei der Wiedergabe ist das Phänomen der Phantomschallquelle.

Strahlen zwei reale Schallquellen kohärente Signale ab, so entsteht eine Phantomschallquelle, da das Gehör die resultierenden Ohrsignale so interpretiert, als würden sie von einer einzigen Schallquelle stammen⁶. Verändert man innerhalb bestimmter Grenzen kontinuierlich die Pegel oder/ und Laufzeitdifferenzen der Signale, so wandert die Phantomschallquelle entlang der Lautsprecherbasis bis sie schließlich in einem der beiden Lautsprecher stehen bleibt (vgl. Dickreiter, Band1, S. 124).

Für die Stabilität der Phantomschallquelle ist es wichtig, dass die beiden Lautsprecher-signale eine möglichst hohe Korrelation aufweisen. Wird der Zeitversatz der beiden Signale zu groß, tritt zuerst das Gesetz der ersten Wellenfront in Kraft und für den Ort des Hörereignisses wird diejenige Schallquelle bestimmend, die das Signal früher abstrahlt. Bei steigender Zeitdifferenz treten schließlich zwei Hörereignisse auf (vgl. Blauert, 1974, S.163). Der Begriff Phantomschallquelle ist meiner Ansicht nach etwas unglücklich gewählt, da dieser suggeriert, das Phänomen der Phantomschallquelle wäre nur ein physikalisches. Es ist jedoch so, dass aufgrund der menschlichen Wahrnehmung ein Hörereignis entsteht, das derart interpretiert wird, als würde es von einer realen Schallquelle stammen.

Blauert definiert in diesem Zusammenhang folgendermaßen: Das *Schallereignis* beschreibt die physikalische Seite des Hörvorgangs, das von einer Schallquelle ausgelöst wird. Als *Hörerereignis* wird das akustisch Wahrgenommene bezeichnet. Das Hörereignis kann durch ein Schallereignis, oder auch durch andere Formen der Reizung des Hörnervs hervorgerufen werden. Auch das Rauschen und Pfeifen bei bestimmten Ohrerkrankungen wird als Hörereignis bezeichnet. Schall- und Hörereignisse treten immer zeitlich, räumlich und eigenschaftlich bestimmt auf. (vgl. Blauert, 1974, S. 2).

Wird daher in den folgenden Ausführungen von der Bewegung der Phantomschallquelle gesprochen, handelt es sich dabei immer um die Interpretation des Menschen, das Hörereignis einer bewegten Schallquelle wahrzunehmen.

⁶ Erklärungen für diese Phänomen liefern u.a. die Summenlokalisierung oder Theiles Assoziationsmodell.

2.6 Downmix-Ansätze

Der automatische Downmix von Mehrkanalaudio kann mit verschiedenen Verfahren in unterschiedlichen Applikationen stattfinden. Neben diverser Hard- und Software-Lösungen ist auch bei der Standardisierung von unterschiedlichen Medien und Audioformaten für Mehrkanalaudio an die Integration von Methoden zum automatischen Downmix gedacht worden.

Ein wesentlicher Unterschied dieser Methoden ist dabei der „Zeitpunkt“ des Downmix. In manchen Medien ist die Möglichkeit des automatischen Downmix als Grundfunktion angelegt und wird unter bestimmten Voraussetzungen nach z.T. vorher festgelegten Parametern ausgeführt. An anderer Stelle wird der Downmix z.B. zum Zwecke der Distribution ganz bewusst mit bestimmten Eigenschaften zusätzlich zum Mehrkanal-audiosignal hergestellt. Der eine Ansatz geht dabei von einem Downmix nur unter bestimmten Voraussetzungen aus, der andere Ansatz liefert gleich zusätzlich zum Mehrkanal-audiosignal einen Downmix.

Letzterer Ansatz spiegelt z.B. die Überlegungen der verantwortlichen Techniker und Ingenieure in den Rundfunkanstalten wieder. Der Downmix ist notwendig, da neben der Mehrkanal-audiofassung zusätzlich ein 2/0-Stereosignal angeboten werden muss, um all diejenigen Empfänger mit Ton zu versorgen, die nicht in der Lage sind, das Mehrkanal-audiosignal wiederzugeben.

Der automatische Downmix ist hier notwendig, da eine zusätzliche Mischung in 2/0-Stereo zu hohe Kosten verursachen würde. So wird der automatische Downmix ergänzend zur Mehrkanalmischung erstellt und in die Sendefunkinfrastruktur integriert. Es ist üblich Mehrkanal-audio und 2/0-Stereo auf zwei unterschiedlichen Distributionswegen zu vertreiben. Der Endverbraucher kann dann auf dem analogen Übertragungsweg den Downmix und auf dem digitalen, z.B. über DVB-T, das Mehrkanal-audiosignal empfangen.

Die zweite Variante findet sich z.B. bei der Dolby Digital Audiospur auf einer DVD-Video wieder. Enthält die Audiospur z.B. einen 3/2/1-Spielfilmsoundtrack im Dolby Digital Format kann der Dolby Digital Decoder die Anzahl der Audiospuren an die Anzahl der zur Verfügung stehenden Lautsprecher anpassen und einen automatischen Downmix der Tonspur durchführen. Der Toningenieur kann dabei im Vorfeld über unterschiedliche Parameter festlegen, auf welche Weise dieser Downmix vollzogen werden soll.

Da die Firma Dolby mit dem Dolby Digital Format und dessen Einbindung in professionelle Produktionsumgebungen hier einige gute Lösungen anbietet, werden im folgenden

Kapitel die Funktionsweise der Dolby Audioformate und die Möglichkeiten, diese für einen automatischen Downmix zu nutzen, näher betrachtet.

3 Dolby Formate

Die Firma Dolby stellt seit vielen Jahren Lösungen für den Audiobereich her und hat mit dem professionellen Format Dolby E in Kombination mit dem Konsumer-Format Dolby Digital eine wesentliche Marktposition bei der Distribution von digitalen Audiosignalen inne. Da sich mit Hilfe der Dolby Formate auch unterschiedliche Downmix-Lösungen anbieten, sollen diese Formate und Downmix-Variationen in diesem Kapitel vorgestellt werden.

3.1 Formate für den Konsumenten

Formate für den Konsumenten werden vornehmlich für die Speicherung, Rundfunkübertragung oder sonstige Distribution verwendet. Neben den analogen Formaten Dolby Pro Logic und Dolby Pro Logic II gehört auch das digitale Format Dolby Digital dazu.

3.1.1 Dolby Pro Logic und Dolby Pro Logic II

Diese Formate ermöglichen es durch eine Matrixierung über jede - analoge oder digitale - zweikanalig ausgelegte Strecke bis zu fünf Kanäle bereitzustellen. Zu den Anwendungsgebieten gehören neben Rundfunk- und Fernsehübertragung auch VHS-Video oder Computer-Spiele.

Dolby Pro Logic wurde 1987 vorgestellt und stellt ein System mit vier Audiokanälen dar: Links, Center, Rechts und ein Mono Surround mit einer begrenzten Bandbreite von 100 Hz bis 7kHz. Ein LFE-Kanal ist nicht vorgesehen. Mittels einer Encoder-Matrix werden diese vier Audiokanäle zu einem 2/0-Stereosignal zusammengemischt, welches von einem Decoder wieder in die ursprünglichen Audiokanäle aufgeteilt werden kann. Ohne einen Decoder wird ein 2/0-Stereosignal aus allen vier Audiokanälen wiedergegeben. Dieses 2/0-Stereosignal wird mit Lt/Rt - „Left Total/ Right Total“ - bezeichnet.

Trotz der neuen qualitativ besseren Audioformate hat der Downmix mit Dolby Pro Logic noch immer eine Bedeutung, da sowohl die DVD-Audio als auch das digitale Tonformat Dolby Digital einen Dolby Pro Logic kompatiblen Downmix als Option anbieten.

Das im Jahre 2000 vorgestellte System *Dolby Pro Logic II* ist eine Weiterentwicklung der Dolby Pro Logic Technik um einen zusätzlichen Surroundkanal. Aufgrund der

zunehmenden Verbreitung von Audioformaten, welche auch dem Heimanwender die Wiedergabe diskreter Kanäle ermöglichen, wurde Dolby Pro Logic II entwickelt, um einen Höreindruck zu ermöglichen, der sich an den neuen diskreten Formaten orientiert.

Im Gegensatz zu Dolby Pro Logic ist die Bandbreite der Surroundkanäle nicht mehr begrenzt und durch die neue, verbesserte Matrix- und Decoder Technik soll die Kanal-trennung verbessert und das Gesamtklangbild stabiler sein.

Dolby Pro Logic II Programme sind vollkommen rückwärtskompatibel und können daher auch auf Dolby Pro Logic, sowie 2/0-Stereo- und Mono-Systemen abgespielt werden. (vgl. Dolby [1], [2], [3])

3.1.1.1 Arbeitsweise des Encoders / Decoders

Der Dolby Pro Logic Encoder

In der Encoder-Matrix wird der Centerkanal um 3 dB abgesenkt und zum linken und rechten Audiokanal addiert. Das Surroundsignal wird in der Phase um 90° gedreht, dann um 3 dB bedämpft und ebenfalls zum linken Audiokanal addiert und vom rechten Audiokanal subtrahiert. Somit liegt das Surroundsignal im Gegensatz zum Centersignal gegenphasig in den beiden Audiokanälen vor. Die Dekodierung wird im Zusammenhang mit dem Dolby Pro Logic II Decoder erläutert.

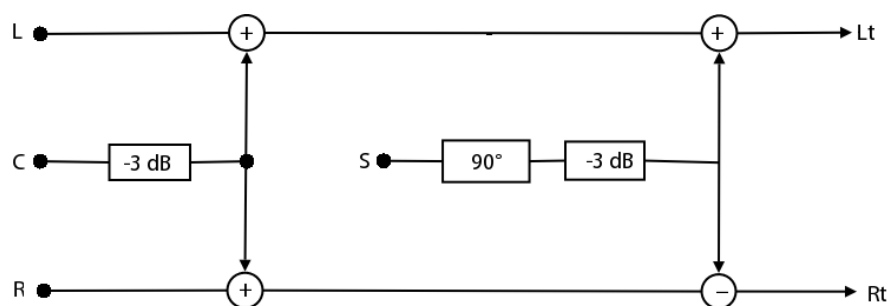


Abbildung 3-1: Dolby Pro Logic Encoder

Wird der Downmix aus einem 3/2-Stereo Dolby Digital Signal erstellt, wird er anhand von vorher festgelegten Parametern durchgeführt, wie beispielsweise nach dem Schema in Abbildung 3-2.

Die beiden Surroundkanäle werden schon bei der Kodierung in das Dolby Digital Format in der Phase um 90° gedreht gespeichert. Bei der Dekodierung, z.B. im DVD-Player werden die Surroundkanäle dann addiert, von 100 Hz bis 7kHz bandpassgefiltert und schließlich als Monosignal zum linken Audiokanal addiert und vom rechten Audiokanal subtrahiert.

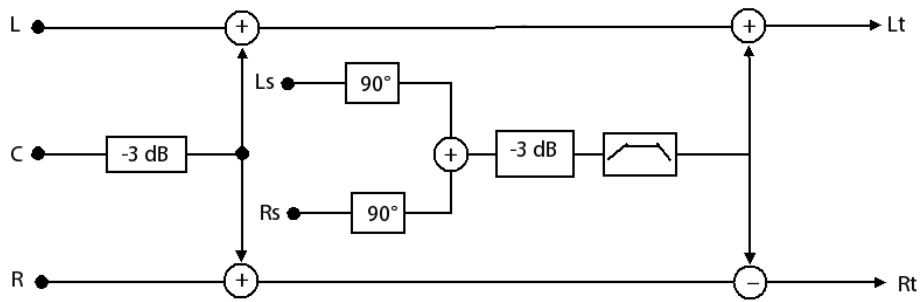


Abbildung 3-2 : Dolby Pro Logic Encoder Variante

Der Dolby Pro Logic II Encoder

Von der Dolby Pro Logic II Encoder Matrix gibt es aufgrund der Firmenpolitik von Dolby nur ein vereinfachtes Blockschaltbild. Die Funktionsweise konnte durch einfache Tests von Florian Rathgeber bestätigt werden (vgl. Rathgeber, S. 38-40). Mit einem Funktionsgenerator wurden Töne mit der Frequenz von 800 Hz generiert und in einer 3/2-Stereo-Mischung auf verschiedene Panorama-Positionen gelegt. Das Ausgangssignal wurde gemessen und mit dem Eingangssignal verglichen.

Der Dolby Pro Logic II Encoder kann bis zu fünf Audiokanäle als Eingangssignale verarbeiten. Der linke und rechte Audiokanal werden direkt auf das Lt/Rt-Signal durchgeschliffen. Der Centerkanal wird um 3 dB abgesenkt und zum linken und rechten Audiokanal addiert. Der linke Surroundkanal wird um 1,2 dB abgesenkt, um 90° in der Phase verschoben und dann vom Lt-Kanal subtrahiert sowie um 5 dB abgesenkt zum Rt-Kanal addiert. Mit dem rechten Surroundkanal wird in gleicher Weise verfahren, wobei das reduzierte Signal auf den Rt-Kanal angewendet wird. Durch die Addition der Surroundkanäle zu beiden Lt/Rt-Stereokanälen wird die Dolby Pro Logic Kompatibilität gewährleistet.

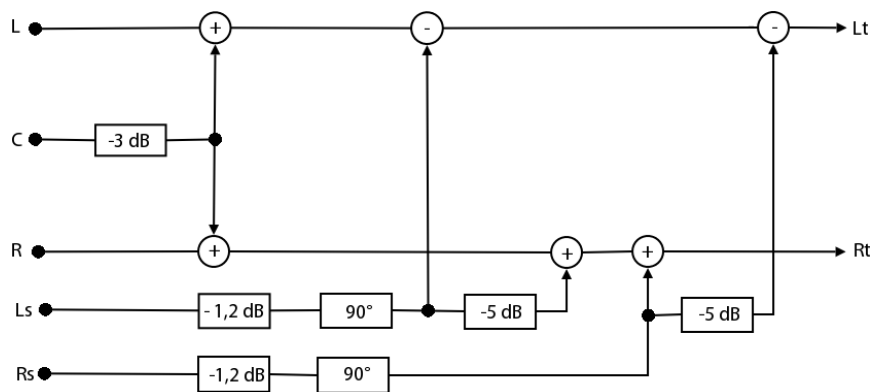


Abbildung 3-3 : Dolby Pro Logic II Encoder Matrix (vgl. Rathgeber, S. 27)

Der Dolby Pro Logic II Decoder

Die bisherigen Dolby-Decoder waren auf dem Prinzip der Vorwärtsregelung aufgebaut. Ein Regelkreis überprüft den relativen Pegel und Phase der Eingangssignale, um diese Informationen an die Ausgangsmatrix zu senden. Dort wird mit Hilfe von spannungsgesteuerten Verstärkern der Pegel der gegenphasigen Signalanteile bestimmt. Durch die gegenphasigen Anteile werden die unerwünschten Übersprechungen auf den einzelnen Audiokanälen ausgelöscht; so führt dieses Verfahren zu einer erhöhten Kanaltrennung.

Der Dolby Pro Logic II Decoder hingegen arbeitet rückgekoppelt. Auch hier werden die Eingangssignale auf relativen Pegel und Phase überprüft. Ein (Servo-) Regelkreis gleicht sie aneinander an und leitet sie direkt zur Matrix weiter, um die einzelnen Ausgangskanäle zu generieren. Außerdem werden die Signale rückgekoppelt auf den Eingang gelegt, um den Regelkreis zu kontrollieren.

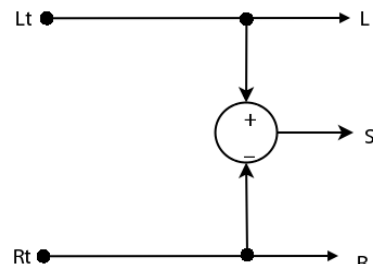


Abbildung 3-4 : Einfacher passiver Surround Decoder (vgl. Dolby® [2], S.3)

Das Surroundsignal liegt nach der Kodierung, wie oben beschrieben, gegenphasig vor. Für den Fall, dass nur ein Surroundsignal am Decodereingang anliegt sind die Audiosignale in Lt und Rt gleich, aber mit entgegengesetzter Polarität. Um die Informationen für die Surroundkanäle zu gewinnen, genügt es den Lt-Kanal vom Rt-Kanal zu subtrahieren. Für die Gewinnung des Centerkanals genügt eine Addition. Die Kanaltrennung funktioniert bei reinen Center- bzw. Surroundsignalen sehr gut. Bei einem reinen Surroundsignal würde durch eine Addition der Lt/Rt-Audiokanäle das Audiosignal ausgelöscht, also im Center kein Signal vorhanden sein.

Problematisch wird es jedoch, wenn gemischte Signale auftreten. Wird z.B. der Dialog zwischen dem Centerkanal und dem rechten Audiokanal positioniert, muss der Decoder aktiv dafür sorgen, dass der Dialog in den beiden Lt/Rt-Kanälen gleichermaßen vorhanden ist. Dieses geschieht mit spannungsgesteuerten

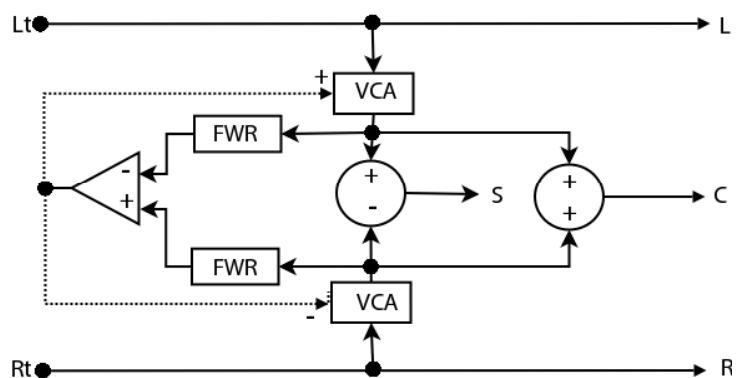


Abbildung 3-5 : VCA- Abgleich im Decoder mit Servo (vgl. Dolby® [2], S. 4)

Verstärkern, die in einer Servo-Schaltung die Kanäle Lt und Rt angleichen. Die dynamische Anpassung der Verstärker geschieht mit Hilfe eines rückgekoppelten Servos, der die Pegel der Verstärker nach einem Vollwellengleichrichter (FWR = full wave rectification) aneinander angleicht. Die Schaltungseinheit, die für eine Kanaltrennung der vorderen Audiokanäle sorgt, nennt man „Links-Rechts-Achse“ des Decoders. Eine zweite Schaltung - „Vorn-Hinten-Achse“ – steuert die Kanaltrennung der vorderen zu den hinteren Audiokanälen. In den Dolby Pro Logic II Decodern arbeiten diese beiden Achsen unabhängig voneinander und passen ihre Geschwindigkeit den jeweils vorherrschenden Bedingungen an.

3.1.1.2 Erstellen einer Lt/Rt-Mischung

Während es bei den alten analogen Encodern für das Dolby Pro Logic Tonformat üblich war, den Abmischvorgang durch eine Encoder-Decoder-Kette abzuhören (vgl. Dolby [4], S. 3-3) und den Mix ggf. durch geschickte Anpassungen an die Encoder Technik zu verbessern, sind die neuen digitalen Encoder für Dolby Pro Logic II darauf ausgelegt eine bestehende 3/2-Stereo-Mischung direkt zu kodieren. Es gilt dennoch einige grundlegende Regeln zu beachten und ggf. das Ergebnis der Kodierung zu kontrollieren.

Zur Optimierung des Downmix empfiehlt es sich den Kodiervorgang am Dolby Pro Logic II Encoder mittels der Parameter „Center Trim“ und „Surround Trim“ an das zu kodierende Material anzupassen: Der Parameter „Center Trim“ sorgt bei Programmen, welche besonders wichtige Informationen im Centerkanal enthalten, für eine Anhebung der Lautstärke des Centerkanals im Ausgang des Encoders. „Surround Trim“ ist für Audiomaterial gedacht, welches besonders viele Informationen in den Surroundkanälen enthält. Durch eine Absenkung des Surround-Pegels lässt sich eine verständlichere Mischung erzielen.

Ein zusätzlicher LFE-Kanal ist für das Dolby Pro Logic II Format nicht vorgesehen. Daher wird von Dolby empfohlen, auf das Hinzufügen des LFE-Kanals zur Lt/Rt-Mischung zu verzichten, es sei denn, der LFE-Kanal enthält Informationen, die in den anderen Kanälen nicht vorhanden sind.

Das aktuelle Modell der Firma Dolby der „Dolby DP563 Dolby Surround und Pro Logic II Encoder“ bietet zudem die Möglichkeit mittels Dolby-Metadaten (siehe unten) den Kodiervorgang anzupassen (vgl. Dolby [3], S.2-3). Dieses Feature sorgt für eine professionelle Einbindung in die Produktionsumgebung.

Der Konsument hat zwei unterschiedliche Einstellmöglichkeiten des Decoders: *Music-Mode* und *Cinema-Mode*.

Der *Music-Mode* ist vor allem in etwas professionelleren Geräten zu finden. Ursprünglich wurde er entwickelt, um Audiomaterial wiederzugeben, welches nicht in Dolby Surround oder Dolby Pro Logic II kodiert wurde. Der Nutzer hat die Möglichkeit, die Informationen des Centerkanals teilweise oder ganz auf den linken und rechten Kanal umzuleiten und über den Parameter „Dimension Control“ das Schallfeld mehr oder weniger stark nach vorne oder hinten zu verschieben. Der „Panorama-Modus“ erweitert die vorderen Audiokanäle auf die Surroundkanäle, um den Eindruck eines den Hörer einhüllenden Klangbildes zu schaffen. Zusätzlich gibt es im Musik-Mode in den Surroundkanälen ein Shelf-Filter, das auf die hohen Frequenzen wirkt. Dadurch soll ein realistischeres Klangbild vermittelt werden, da der natürliche Nachhall aufgrund von Absorption und Reflexion eine leichte Höhendämpfung aufweist (vgl. Dolby [2]).

Die meisten Decoder verfügen jedoch nicht über die Wahlmöglichkeit zwischen alternativen Einstellungen. Hier ist der *Cinema-Mode* bestimmend für die Arbeitsweise des Decoders. Der Cinema-Mode ist dem Dolby Pro Logic Decoder sehr ähnlich, wobei der Cinema-Mode mit zwei nicht bandbegrenzten Surroundkanälen arbeitet. Im Cinema-Mode, sowie im „Dolby Pro Logic Emulations Modus“ werden die Surroundkanäle zeitverzögert wiedergeben. So wird im Decoder der Haas-Effekt ausgenutzt, um eine bessere Klangwiedergabe und Ortung zu gewährleisten.

Nach der Enkodierung von Mehrkanal-Audiosignalen in das Dolby Pro Logic II Format sollte auf eine weitere Bearbeitung mit Signalprozessoren verzichtet werden, da die meisten dieser Prozesse zu Phasenverschiebungen führen können, welche die Matrixkodierung unbrauchbar machen (vgl. Dolby [1], S. 6). Alle anderen Prozesse sollten auf beide Audiokanäle exakt gleich angewendet werden. Dies gilt vor allem für Kompressoren, Limiter oder Delays.

(vgl. Dolby [1], Dolby [2], Dolby [3], Dolby [4])

3.1.2 Dolby Digital

Der Dolby Digital Codec hat sich mittlerweile als „Quasi Standard“ für den Endbenutzer etabliert. So ist er einer der nicht-optionalen Audiocodecs der DVD-Video und wurde auch für das digitale Fernsehen spezifiziert.

Dolby Digital basiert auf dem Modell der digitalen Audio-Komprimierung AC-3⁷. Der AC-3-Codec ist ein sogenannter „Perceptual Codec“ und beruht auf Eigenschaften der menschlichen Wahrnehmung. Der Codec ist in der Lage bis zu sechs diskrete Audiokanäle digital mit verlustbehafteter Datenreduktion zu übertragen. Dabei kann es sich um fünf Audiokanäle mit einem Frequenzumfang von 20 Hz bis 20 kHz und einen optionalen LFE-Kanal bis 120 Hz handeln. Die unterstützten Bitraten liegen zwischen 32 bis 640 kbps; auf einer DVD werden z.B. für 3/2/1-Stereo Bitraten von 384 bis 448 kbps verwendet, für 2/0-Stereo liegt die Bitrate zwischen 192 und 224 kbps.

Dolby Digital ist jedoch weitaus mehr als ein Komprimierungs-Codec. Es gibt dem Produzenten außerdem die Möglichkeit, über sogenannte Metadaten festzulegen, in welcher Weise das Programm beim Nutzer zu hören sein wird. Diese Daten können z.B. die relative Wiedergabelautstärke beeinflussen oder festlegen, auf welche Art ein Downmix im Decoder erfolgen soll.

3.1.2.1 Funktionsweise des Encoders

Ziel der Kodierung ist es, aus dem digitalisierten Audiosignal Bits zu entfernen, so dass im Endergebnis kein Unterschied zum Original hörbar wird.

Das Fundament eines jeden Audio-Encoders ist die sogenannte „Bit Allocation“ (Zuteilung des „Bit- Budgets“). Es gibt zwei Klassen von Strategien, die hier Verwendung finden: „forward adaptive“ und „backward adaptive“ Bit Allocation.

Forward Bit Allocation bedeutet, dass der Encoder die Neuzuteilung der Bits berechnet und Informationen über die angewandte Methode explizit in den Datenstrom schreibt.

Backward hingegen bezieht Informationen über die Neuzuteilung aus den übertragenen Audiodaten. Da beide Methoden verschiedene Vor- und Nachteile haben, verwendet AC-3 eine sogenannte „Hybrid Backward/ Forward Adaptive Bit Allocation“, mit der die meisten der Nachteile umgangen werden.

Der Encoder speichert die PCM-Eingangswerte zunächst in einem Eingangspuffer und arbeitet dann mit Blöcken von 512 Samples (siehe Bild 3-7[a]). Mit Hilfe eines Hochpasses mit einer Grenzfrequenz von 3 Hz werden Gleichanteile aus dem Signal entfernt (siehe Bild 3-7 [b]). Bei einem evtl. vorhandenem LFE-Signal werden alle Signalanteile oberhalb von 120 Hz herausgefiltert.

⁷ www.atsc.org - A/52 – Digital Audio Compression Standard (AC-3)
A/54 – Guide to the Use of ATSC Digital Television Standard

Nun folgt eine Umwandlung der Audio Blöcke vom Zeit- in den Frequenzbereich. AC-3 bedient sich hierzu der TDAC-Filter-Bank (siehe Bild 3-7 [d]), wie sie von Princen und Bradley⁸ beschrieben wird: Nach der Fensterung mit einer Besselfunktion, wird mittels einer FFT eine MDCT (modifizierte Diskrete-Kosinus-Transformation) durchgeführt, um die Audio-Blöcke in den Frequenzbereich zu überführen. Die Blöcke überschneiden sich jeweils um 50 %; da somit jeder Zeitwert von zwei Frequenzwerten repräsentiert wird, werden aus 512 Samples im Zeitbereich 256 Frequenzwerte. Bei einer Abtastrate von 48 kHz repräsentiert dabei ein Block jeweils $\frac{512\text{Samples}}{48\text{kHz}} = 10,66\text{ms}$ Audiomaterial. Aus den doppelt vorhandenen Samples resultiert eine Transformation für alle 5,33 ms (=Blocklänge) und eine Blockrate von $\frac{48\text{kHz}}{256} = 187,5\text{Hz}$ pro Audio-Block.

Während stationäre Audiosignale von einer hohen Frequenzauflösung profitieren, sind Transienten mit einer hohen zeitlichen Auflösung besser zu beschreiben. Wenn Transienten zeitlich zu gering aufgelöst werden, können sie zu Quantisierungsrauschen führen, da sie über die Zeit „verschmiert“ erscheinen. Da es nicht möglich ist, im Zeitbereich sowie im Frequenzbereich gleichermaßen fein aufgelöst zu arbeiten, musste ein Kompromiss gefunden werden.

Für Transienten wird mit Hilfe eines Transienten-Detektors (siehe Bild 3-7 [c]) - einem hochfrequenter Bandpassfilter - die Zeitauflösung erhöht, indem die Blockgröße halbiert wird, so dass nun alle 2,67 ms eine Transformation stattfindet, mit einer Blockrate von 93,75 Hz.

Damit stehen auch die minimalste Zeitauflösung und Blockrate für das TDCA-Filter fest. Das Filter ist zwischen 0 und 24 kHz in 50 Frequenzbänder unterteilt, mit einer Bandbreite zwischen $\frac{3}{4}$ und $\frac{1}{4}$ der „critical bands“⁹.

Jeder einzelne Frequenz-Koeffizient wird nun im Floating-Point-Format (mit einer Bittiefe von 16 bis 18 Bit; maximal 24 Bit) kodiert (siehe Bild 3-7 [e]). Die Mantisse umfasst dabei maximal 16 Bit und der Exponent 5 Bit.

⁸ Princen, J.; and A. Bradley, "Analysis/ Synthesis Filter Band Design Based on Time-Domain Aliasing Cancellation," IEEE Trans. ASSP, vol. ASSP- 34, no. 5, October, 1986 (vgl. Pohlmann, S. 697)

⁹ Man geht davon aus, dass das Gehör Schallereignisse in Frequenzgruppen auswertet. Es werden dafür fünf Frequenzgruppen mit je 100 Hz Bandbreite für Frequenzen kleiner als 500 Hz und 19 Frequenzgruppen mit einer Bandbreite von 20 % der Mittenfrequenz für Frequenzen größer als 500 Hz angenommen. Im englischsprachigen Bereich werden diese Frequenzgruppen als „critical bands“ bezeichnet. (vgl. Adam, S. 22).

Die Exponenten werden innerhalb eines Audio-Blocks in Gruppen („Spectral Envelope“) zusammengefasst (siehe Bild 3-7 [f]) und von nun an getrennt von den Mantissen bearbeitet. Dazu werden drei verschiedene Methoden – D15, D25, D45 – angewandt, die als „Exponent Strategies“ bezeichnet werden. Bei allen Methoden wird eine gewisse Anzahl an Exponenten zu 7 Bit Wörtern zusammengefasst. Der Hauptunterschied liegt dabei in der Anzahl der verwendeten Exponenten. Die Exponenten werden zudem nicht mit Absolutwerten, sondern differentiell kodiert.

Die Filter der „Transform Filter Bank“ weisen eine Steigung von 12 dB pro Oktave auf. Diese Tatsache wird ausgenutzt, um die Kodierung zu optimieren. Da die Werte von zwei benachbarten Filtern selten mehr als eine Differenz von 12 dB aufweisen, können Deltas mit einem Umfang von 6 dB gebildet und durch 5 einfache Zustandsänderungen (+2, +1, 0, -1, -2) ausgedrückt werden. Die Exponenten werden nun codiert, indem der erste Exponent einer Gruppe einen Absolut-Wert erhält, die folgenden Werte die Abweichung zu diesem. In einem 7 Bit Wort werden immer drei dieser Differentiale zusammengefasst.

D15 speichert für jeden abweichenden Frequenzwert ein Delta. Daraus ergibt sich eine Datenrate von 2,33 bit pro Exponent. Diese Methode bietet eine hohe Auflösung im Frequenzbereich auf Kosten der zeitlichen Auflösung und wird angewandt, wenn die Envelope über viele Blöcke konstant bleibt. Da das Audiosignal relativ konstant ist, wird nicht für jeden Block ein Wort gespeichert, sondern z.B. alle sechs Blöcke (Datenrate $2,33/6 = 0,39$ bit/Exponent).

Bei einem weniger konstanten Audiosignal werden die anderen beiden Methoden verwendet: *D25* ergibt ein ausgeglichenes Verhältnis von Zeit- und Frequenzauflösung und wird angewendet, wenn das Spektrum über zwei bis drei Audio-Blocks relativ stabil bleibt. Zwei Exponenten bilden dabei ein Differential. Die Datenrate beträgt 1,16 bit pro Exponent (die Hälfte von *D15*) und hat eine um den Faktor zwei schlechtere Frequenzauflösung als *D15*.

Bei der Methode *D45* wird alle vier Exponenten ein Delta gespeichert, diese Methode beansprucht nur ein viertel der Datenrate von *D15*; es ergibt sich eine Datenrate von 0,58 bit pro Exponent. Man erhält dabei eine hohe Auflösung im Zeitbereich auf Kosten der Auflösung im Spektrum. *D45* wird benutzt, wenn Transienten in den einzelnen Audio-Blöcken auftreten ([A/52B], S. 56- 57).

Der Encoder wählt eine dieser drei Methoden oder REUSE (Wiederverwenden) aus und speichert in einem 2 Bit Flag, welche Methode vom Decoder angewendet werden soll.

Der nächste Bearbeitungsblock ist die „Global Bit Allocation“ (Bild 3-7 [g]), also die Zuweisung der benötigten Bits für jede Mantisse. Die TDCA-Koeffizienten werden auf ihre Maskierungseigenschaften (auch unterhalb der einzelnen Audiokanäle) bezüglich des menschlichen Hörapparates untersucht. Sie werden dazu mit einer Funktion gefaltet, die den Maskierungseigenschaften des menschlichen Hörapparates entspricht.

Um den Rechenaufwand so gering wie möglich zu halten, wird dabei die Bandbreite des Spektrums verringert. Je nach Bandbreite des erwünschten Signals werden bis zu 252 Frequenzwerte auf 64 Subbänder aufgeteilt. Proportional zu den „critical bands“ wächst die Bandbreite von 1 bei tiefen Frequenzen auf 16 bei hohen Frequenzen.

Um die eigentliche Faltung durchzuführen, wird ein vereinfachtes Verfahren durchgeführt. Die Maskierungsfunktion wird durch zwei Kurven angenähert: Eine sehr schnell und eine langsam abfallende Maskierungs-Kurve. Für jede der Kurven („slow leak“ und „fast leak“) wird beginnend bei der tiefsten Frequenz berechnet, ob die Werte für den Gesamtklang relevant sind oder nicht. Das Ergebnis wird mit der Hörschwelle verglichen und vom Originalsignal subtrahiert, um die „Signal to Noise Ratio“ (SNR) für die einzelnen Koeffizienten zu ermitteln. Hiermit wird dann jede einzelne Mantisse der Koeffizienten quantisiert (siehe Bild 3-7 [h]).

Die zur Verfügung stehenden Bits für alle Kanäle stammen aus einem gemeinsamen Bit-Pool. Zum Schluss wird die Anzahl der verbrauchten Bits errechnet; wenn noch Bits zur Verfügung stehen, wird die SNR der Mantisse erhöht; wenn die Anzahl der Bits überschritten wurde muss die SNR verringert werden (dann wird an dieser Stelle auch gedithert) oder es wird die Technik des „Couplings“ angewendet.

„Coupling“ nutzt die Richtungswahrnehmung von hohen Frequenzen aus: Bei hohen Frequenzen (ab ca. 1,6 kHz) ist das Ohr nicht mehr in der Lage, die einzelnen Wellen eines Audiosignals aufzulösen, sondern die Richtung wird anhand der interauralen Zeitverzögerung der Hüllkurven (Enveloppen) bestimmt (vgl. Blauert, 1974, S.122). Coupling nutzt diesen Effekt aus. Wenn also das „Bit-Budget“ knapp wird, werden verschiedene Kanäle über ihre Eigenschaften der hohen Frequenzen zusammengefasst. Die Coupling-Frequenz bezeichnet dabei die Grenze, ab welcher Frequenz Coupling angewandt wird. Oberhalb dieser Frequenz werden die einzelnen Kanäle zu einem Coupling-Channel zusammengefasst, wobei auf die Phase der einzelnen Kanäle geachtet

werden muss, um Auslöschungen zu vermeiden. Der Encoder generiert *Coupling-Koordinaten* für jeden individuellen Kanal, um das ursprüngliche Signal wieder herzustellen. Die Coupling-Channel werden genau wie die übrigen Kanäle mit Envelope aus Exponent und quantisierter Mantisse übertragen. Der Encoder bestimmt die Coupling-Frequenz und welche Audiokanäle zusammengefasst werden. Die jeweils angewandte Methode liegt dabei im AC-3 Bitstream vor, so dass das Coupling individuell angepasst werden kann. Bei den heutigen AC-3 Encodern liegt die Coupling-Frequenz durchschnittlich bei 10 kHz.

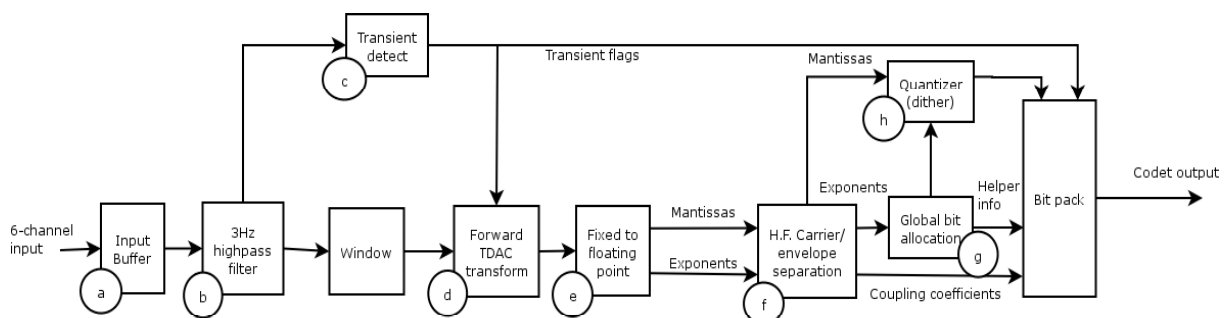


Abbildung 3-7 : AC- 3 Encoder (vgl. Dolby [8], S. 6)

Als letzter Schritt werden alle Daten zum AC-3 Datenstrom zusammengefügt. Die AC-3 Daten liegen in Sync-Frames vor, wobei jeder Frame eine unabhängige Einheit darstellt. Ein Frame enthält einen Synchronisations-Information-Header (SI-Header), einen Bitstream-Information-Header (BSI), 32 ms Audiodaten in Form von quantisierten Frequenz-Koeffizienten, ein Auxiliary Feld und CRCC-Fehlerschutzdaten¹⁰.

Der SI-Header besteht aus einem 16 Bit Sync-Wort gefolgt von 8 Bit Informationen über Sample-Rate (2 Bit) und Frame-Größe (6 Bit). Das BSI-Feld beschreibt die Audiodaten mit Informationen über Kodierung- Modus, Timecode, Copyright, Dialog-Level und Language Code. Es folgen 6 Audio-Blöcke wie oben beschrieben.

Unbenutzte Daten werden im Auxiliary-Feld zusammengefasst. Am Ende jedes Frames folgt ein 16 Bit langes CRCC-Wort. Optional kann auch ein 16 Bit CRCC-Wort im SI-Header platziert sein.

Einer oder mehrere AC-3 Datenströme können in einem MPEG-2 Transport Strom eingebettet sein, wie in der ISO/ IEC 13818-1 beschrieben.

¹⁰ CRCC-Fehlerschutz beruht auf der Polynomen-Division. Vor Beginn der Übertragung wird ein CRCC-Wort berechnet. Nach der Übertragung wird das CRCC-Wort erneut berechnet und beiden Prüfwerte verglichen. Das Polynom bei AC-3 ist $x^{16} + x^{15} + x^2 + 1$ (vgl. Pohlmann, S. 137– 142, S. 355).

3.1.2.2 Der AC-3 Decoder

Der Decoder speichert genau wie der Encoder am Eingang Code in einem Buffer, um diesen dann blöckeweise zu bearbeiten (siehe Bild 3-8 [a]). Daraufhin folgt eine Fehlerkorrektur (siehe Bild 3-8 [b]). Falls ein nicht zu korrigierender Fehler auftritt, verwendet der Decoder den letzten richtigen Block. Durch das Überschneiden der einzelnen Blöcke ist eine wiederholte Verwendung der Blöcke gut möglich, bevor der Decoder das Audiosignal unterbrechen, oder - wie bei Kinofilmen üblich - auf eine optionale, analoge Tonspur umschalten muss.

Die nächste Stufe ist das „Fixed format data unpacking“ (Bild 3-8 [c]). Dieses betrifft die Exponenten, Coupling-Koeffizienten und diverse „Mode-Flags“. Mit diesen Daten wird nun die „Global Bit Allocation“ (siehe Bild 3-8 [d]) nachvollzogen, um die variablen Daten zurückzugewinnen (siehe Bild 3-8 [e]). Falls erforderlich wird ein „Decoupling“ vollzogen, um die hochfrequenten Anteile der einzelnen Kanäle wieder einzufügen und die Enveloppen den richtigen Frequenzen zuzuordnen (siehe Bild 3-8 [f]). Optional werden an dieser Stelle die Dynamik-Range-Compression (siehe unten) und andere zusätzliche Prozesse auf den Audio-Block angewendet. Als Vorbereitung der inversen TDCA-Transformation werden nun die einzelnen Koeffizienten wieder in das Fixed-Point-Format zurückgewandelt (siehe Bild 3-8 [g]) und evtl. vorhandener Dither entfernt. Die inverse TDCA-Transformation der Daten (siehe Bild 3-8 [h]) zurück in die Zeitebene, sowie Fensterung und Operationen, welche das Überlappen der einzelnen Blöcke betreffen, werden durchgeführt, bevor das Signal nach Bufferung als kontinuierlicher PCM-Datenstrom vorliegt.

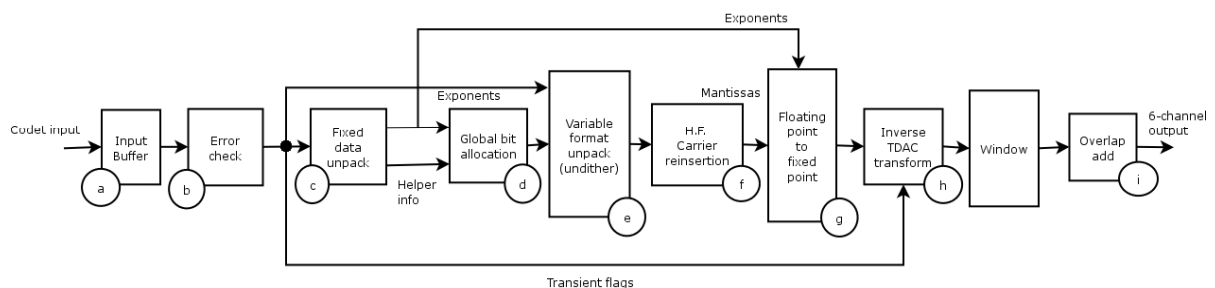


Abbildung 3-8 : AC- 3 Decoder (vgl. Dolby [8], S. 6)

Bei vorhanden LFE-Kanal werden vor der Transformation die Bereiche der mittleren und hohen Frequenzen mit Nullen aufgefüllt, um zu gewährleisten, dass die Sample-Rate kompatibel zu den übrigen Kanälen ist. (vgl. Pohlmann, S. 348- 355; Dolby [7], Dolby [8])

3.1.2.3 Metadaten

Metadaten sind Kontroll-Parameter, mit deren Hilfe sich verschiedene Zusatzinformationen im Dolby Digital-Signal übertragen lassen. Sie dienen dazu, die Wiedergabelautstärke zu beeinflussen, die Art eines Downmix festzulegen oder um programmspezifische Parameter zu übertragen. Die wichtigsten dieser Parameter sollen hier kurz erläutert werden.

Audio Coding Mode

Der Parameter *Audio Coding Mode* oder *Channel Mode* gibt Auskunft über die im Datenstrom vorhandenen Audiokanäle mit vollem Frequenzumfang sowie ihre Zuordnung und Datenraten.

| Audio Coding Mode | Kanal Format |
|--------------------------|---------------------|
| 1+1 | L/Ch1, R/Ch2 |
| 1/0 | C |
| 2/0 | L, R |
| 3/0 | L, C, R |
| 2/1 | L, R, S |
| 3/1 | L, C, R, S |
| 2/2 | L, R, LS, RS |
| 3/2 | L, C, R, LS, Rs |

Tabelle 3-1: Audiocoding Mode

Dolby Digital kann Datenraten von 56 kb/s bis 640 kb/s zur Verfügung stellen. Je nach Anwendungsfall werden von Dolby folgende Datenraten empfohlen:

| Audio Coding Mode | Mögliche Datenrate | Empfohlene Datenrate |
|--------------------------|---------------------------|-----------------------------|
| 1/0 | 56-640 kb/s | 96 kb/s |
| 2/0 oder 1+1 | 96-640 kb/s | 192 kb/s |
| 3/1 oder 2/2 | 192-640 kb/s | 384 kb/s |
| 3/2 | 224-640 kb/s | 448 kb/s |

Tabelle 3-2: Audiocoding Mode Datenraten

Dialog Normalisation

Dialog Normalisation dient dazu, die Lautstärke von unterschiedlichem Audiomaterial dynamisch anzupassen, so dass der Hörer eine einheitliche Gesamtlautstärke erwarten kann, wenn er zwischen verschiedenen Programmen hin- und herwechselt. Wenn der Hörer z.B. von den Nachrichten, die er über sein DVB-T Gerät empfängt, zu einem Film auf seinem DVD-Spieler wechselt, erkennt der Dolby Digital Decoder anhand des vom Produzenten vorher festgelegten *dialnorm*- oder *Dialog-Level*-Wertes, ob die Gesamtlautstärke gesenkt oder angehoben werden muss.

Der Dialog-Level-Wert beschreibt eine Durchschnittslautstärke, welche der Dialog in einem Programm über längere Zeit erreicht. Der Referenzwert liegt hierbei bei -31 dBFS des digitale Ausgangs.

Der Dialog-Level-Wert ist von -1 dB bis -31 dB in 1 dB Schritten einstellbar; es gilt:

$$31 + \text{Dialog-Level-Wert} = \text{Lautstärke-Anpassung}^{11}.$$

Für die Einstellung des Dialog-Level-Wertes gibt es keine klaren Vorgaben; die Werte sind immer abhängig vom jeweiligen Programm. Generell ist ein Wert von -31 dB eher ungewöhnlich und wird am ehesten bei Film-Soundtracks mit einem sehr großen Dynamikumfang Anwendung finden. Für gewöhnlich ist ein Wert zwischen -15 dB bis -20 dB empfehlenswert. Zur Bestimmung des Dialog-Level-Wertes wird von Dolby ein Verfahren aus der Schallimmissions-Messung zu Grunde gelegt, welche mit $L_{eq}(A)$ bezeichnet wird. Dabei wird der energieäquivalente, A-bewertete Schalldruckpegel über eine bestimmte Zeit integriert und zum Signalpegel von 0 dBFS in Bezug gesetzt. Das „Broadcast Loudness Meter LM100“ von Dolby verwendet dabei entweder eine Integration über 10 Sekunden oder eine Langzeitmessung über den gesamten Betrachtungszeitraum. Auch die Firma RTW stellt mit dem „RTW Surround Monitor 10800X“ ein Gerät zu Verfügung mit dem sich der Dialog-Level-Wert bestimmen lässt (vgl. Steveaux, S.25, S.47). Zur Orientierung kann man ebenfalls die von Dolby herausgegebene „Dolby Digital Dialog Normalization Disc“ verwenden, auf der sich einige Hörbeispiele befinden.

¹¹ Um beispielsweise das Audiomaterial insgesamt um 10 dB anzuheben, ist ein „dialnorm“- Wert von -21 dB einzustellen; ein Wert von -31 dB entspricht also keiner Änderung.

Dynamic Range Control

Mit Hilfe des Parameters *Dynamic Range Control* lässt sich die Dynamik des Programms anpassen, um das Hörerlebnis zu verbessern. Obwohl es mit Dolby Digital möglich ist, dass Audiomaterial mit dem selben Dynamikumfang abzuhören, den der Produzent vorgesehen hat, gibt es dennoch Gründe diesen einzuschränken. Wenn man z.B. in seinem Wohnzimmer einen Film mit Dolby Digital Ton genießt, hat man die Gesamtlautstärke seines Verstärkers so eingestellt, dass die Dialoge der Schauspieler gut zu verstehen sind. Sehr laute Geräusche, wie z.B. Explosionen können jedoch so laut werden, dass sie für den „Heimgebrauch“ inakzeptabel sind. Ein anderes Beispiel wäre ein Sinfoniekonzert: Bei einer angenehmen Lautstärke für die lauten Passagen des Werkes könnten die leisen Passagen durch Hintergrundgeräusche verdeckt werden.

Dynamic Range Control erlaubt daher eine Einschränkung der Dynamik. Zusätzlich bietet der Encoder einen Limiter zum Schutz vor Verzerrungen durch zu hohe Pegel.

Der Dolby Digital Encoder generiert Kontrollparameter für die Kompression des Audiomaterials, die im Dolby Digital Datenstrom enthalten sind. Die Parameter sind abhängig von den jeweiligen Einstellungen der Parameter für die Dialog Normalisation, dem gewünschten Dynamic Range Compression Profil und dem zu kodierenden Audiomaterial. Es gibt sechs unterschiedliche Dynamic Range Compression Profile, die später beschrieben werden sollen. Für jedes dieser Profile gibt es fünf Regionen, in denen das Audiosignal auf unterschiedliche Weise komprimiert wird. Sie alle liegen um ein „Null-Band“, das relativ zum gewählten Dialog-Level-Wert liegt.

- Constant Boost → sehr leise Signale
- Variable Boost → leise Signale
- Keine Kompression (Null-Band)
- Variable Cut: laute Signale
- Constant Cut: sehr laute Signale

Diese fünf Regionen werden durch die Parameter Maximum Cut, Maximum Boost, Cut Ratio, Boost Ratio, Breite und Lage des Null Bandes festgelegt (siehe Abbildung: 3-9).

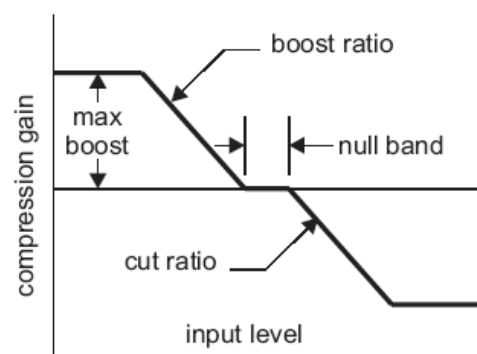


Abbildung 3-9: Dynamic Range Control (vgl. Dolby® [5], S. C-3)

Der Produzent kann nun aus fünf Dynamic Range Compression Profilen wählen, welches Profil bei aktivierten Dynamic Range Control am Decoder ausgeführt werden soll. Abhängig von der Auswahl ändern sich für jedes Profil die Parameter der Regionen (siehe Tabelle 3-3).

| Profile Name | Film Standard | Film Light | Music Standard | Music Light | Speech |
|--------------------------------|--------------------------|-----------------------|---------------------------|------------------------|-----------------------|
| time constant (tc) selection | | | | | |
| attack threshold | 15 dB | 15 dB | 15 dB | 15 dB | 10 dB |
| decay threshold | 20 dB | 20 dB | 20 dB | 20 dB | 10 dB |
| fast attack tc | 10 ms | 10 ms | 10 ms | 10 ms | 10 ms |
| slow attack tc | 100 ms | 100 ms | 100 ms | 100 ms | 100 ms |
| slow decay tc | 3 s | 3 s | 10 s | 3 s | 1 s |
| fast decay tc | 1 s | 1 s | 1 s | 1 s | 200ms |
| hold off period | 10 blocks (53 ms) | 10 blocks (53 ms) | 10 blocks (53 ms) | 10 blocks (53 ms) | 10 blocks (53 ms) |
| Compression characteristic | | | | | |
| max boost (abs range) | 6 dB (-43 dBFS) | 6 dB (-53 dBFS) | 12 dB (-55 dBFS) | 12 dB (-65 dBFS) | 15 dB (-50 dBFS) |
| boost ratio (abs range) | 2:1 (-43 to -31) | 2:1 (-53 to -41) | 2:1 (-55 to -31) | 2:1 (-65 to -41) | 5:1 (-50 to -31) |
| null band width (abs range) | 10 dB (-31 to -21) | 20 dB (-41 to -21) | 10 dB (-31 to -21) | 20 dB (-41 to -21) | 10 dB (-31 to -21) |
| cut ratio (abs range) | 20:1 (-21 to +4) | 20:1 (-21 to +4) | 20:1 (-21 to +4) | 2:1 (-21 to +9) | 20:1 (-21 to +4) |
| max cut (abs range) | 24 dB (+4 dBFS) | 24 dB (+4 dBFS) | 24 dB (+4 dBFS) | 15 dB (+9 dBFS) | 24 dB (+4 dBFS) |

Tabelle 3-3 : Dynamic Range Compression Profiles Parameter für den Line Mode (vgl. Dolby [5], S. C-4)

Ein Schutz vor digitalem Clipping durch den Downmix geschieht während der Kodierung. Der Encoder generiert mehrere mögliche Varianten des Downmix in Abhängigkeit des Parameters für die Dialog Normalisierung und berechnet daraus, um welchen Pegel das Audiomaterial reduziert werden muss.

Einige Dolby Digital Decoder bieten zwei unterschiedliche Audioausgänge an: Einen Line Ausgang (z.B. zum Anschluss an den Verstärker der HiFi-Anlage) und einen Antennenausgang um TV-Geräte oder VHS-Recorder über das Antennenkabel zu speisen. Für beide Ausgänge bietet der Dolby Digital Decoder unterschiedliche Betriebsmodi: Den „Line Mode“ und den „RF Mode“. Für beide Modi ist die Arbeitsweise des Decoders bzgl. der Dynamic Range Control Parameter unterschiedlich.

Im *RF Mode* wird im Gegensatz zum Line Mode das gesamte Programm mit einer Lautstärkeanhebung von +11 dB, also bei -20 dBFS wiedergegeben. Im Falle eines Downmix wird jeder Kanal um 3 dB abgesenkt. Außerdem kann der Encoder auch die Preemphase beim Downmix berücksichtigen, die von einem RF Modulator erzeugt wird um die Gesamtlautstärke anzupassen¹².

Downmix

Für den Fall, dass den Benutzer weniger Abhörwege zur Verfügung stehen, als im Dolby Digital Datenstrom vorhanden sind, kann der Decoder einen automatischen Downmix der Kanäle erstellen (siehe Abbildung 3-10)¹³.

Zusammenfassend sollen hier diejenigen Parameter angesprochen werden, die für einen Downmix mit dem Dolby Digital Decoder Bedeutung haben.

Der Parameter *Center Downmix Level* gibt die Lautstärke an, mit welcher der Centerkanal im Falle eines Downmix zum rechten und linken Audiokanal addiert wird. Es sind drei Lautstärkeabstufungen möglich: -3 dB (Standart), -4,5 dB und -6 dB.

Surround Downmix Level hat die gleiche Funktion wie Center Downmix Level, bezogen auf die Surroundkanäle. Auch hier sind drei Abstufungen möglich: -3 dB (Standart), -6 dB oder „ausgeschaltet“.

¹² Die Berücksichtigung der Preemphase geschieht mit dem Parameter *RF Overmodulation Protection*. Dolby® empfiehlt allerdings diesen Parameter nur in seltenen Fällen einzuschalten (vgl. Dolby® [6], S. 17).

¹³ Einige DVD-Spieler bieten zusätzlich zu den digitalen Ausgängen, analoge Stereo-Ausgänge. An diesen Ausgängen liegt je nach Einstellung der Downmix-Parameter im Decoder entweder ein Dolby® Surround kompatibles Lt/Rt- oder ein Lo/Ro-Audiosignal nach ITU (siehe Kapitel 4.1) an. Empfänger für digitales Fernsehen unterstützen erst mit Einführung von HDTV den Dolby® Pro Logic kompatiblen Downmix. Bis dahin liegt an den analogen Ausgängen MPEG-Stereo.

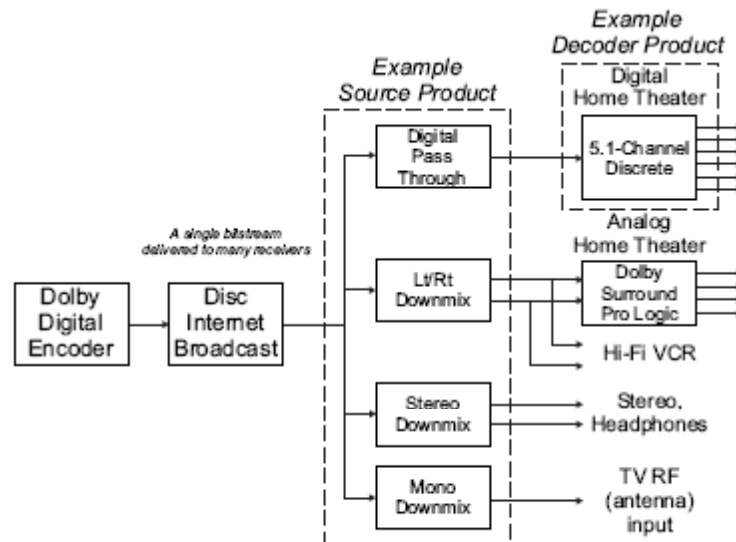


Abbildung 3-10 : Wiedergabemöglichkeiten von Dolby Digital (vgl. Dolby [5], S.3-4).

Surround Phase Shift ermöglicht es dem Dolby Digital Decoder einen zu Dolby Surround kompatiblen Lt/Rt-Downmix zu erstellen. Dazu wird die Phase der Surroundkanäle bei der Kodierung um 90° gedreht. Für die meisten Audiosignale ist dieser Effekt bei der Decodierung zu einem diskreten 3/2-Stereo-Signal nicht von großer Bedeutung, und Dolby empfiehlt diesen Parameter zu setzen. Bei Audiomaterial, das sehr von der Phasenlage abhängig ist (z.B. Musik), sollte dennoch die Notwendigkeit eines Lt/Rt-Downmixes gegenüber der Klangqualität der diskreten 3/2-Stereo-Mischung abgewogen werden¹⁴.

Zusätzlich zu diesen Standard-Parametern (Bitstream Information Parameter) sind einige Decoder in der Lage, eine neue, zusätzliche Generation von Informationen aus dem Dolby Digital Datenstrom auszulesen; Die *Extended Bitstream Information Parameter*:

Mit *Preferred Stereo Downmix Mode* kann der Produzent festlegen, ob aus den analogen 2/0-Stereo Ausgängen eines Konsumer-Decoders ein Lt/Rt- oder Lo/Ro-Downmix (siehe Kapitel 4.1) bevorzugt ausgegeben werden soll. Diese Eingabe kann jedoch vom Endbenutzer mit den Einstellungen am Decoder „überschrieben“ werden.

Lt/Rt Center Downmix Level und *Lt/Rt Surround Downmix Level* für einen Lt/Rt-Downmix, sowie *Lo/Ro Center Downmix Level* und *Lo/Ro Surround Downmix Level* für einen Lo/Ro-Downmix arbeiten genau wie die Einstellungen *Center Downmix Level* und *Surround Downmix Level*, haben jedoch Priorität, wenn der Decoder in der Lage ist, diese

¹⁴ Da die Beeinflussung der Phasenlage während der Kodierung in das Dolby® Digital Format geschieht, ist diese somit auch im dekodierten Audiomaterial noch vorhanden. Die Entscheidung, die Beeinflussung der Phasenlage nicht bei der Decodierung durchzuführen, wurde zum Zeitpunkt der Entwicklung aus wirtschaftlichen Gründen getroffen, da die Decoder sonst zu teuer gewesen wären.

Informationen auszuwerten. Zudem sind die Einstellmöglichkeiten der Lautstärke verfeinert worden: Einstellungen sind in 1,5 dB Schritten von +3 dB bis -6 dB möglich; dazu kommt die Möglichkeit Center und Surroundkanäle auszuschalten.

Sonstige Parameter

Diese Zusatz-Parameter werden an dieser Stelle kurz erwähnt, ohne jedoch im Detail erläutert zu werden.

Dolby Surround Mode kann gesetzt werden, wenn es sich bei dem 2/0-Stereosignal des Audio Coding Modes 2/0 um ein Dolby Surround kompatibles Lt/Rt-Signal handelt.

LFE Channel zeigt in Abhängigkeit vom Channel Mode an, ob ein LFE-Kanal im Datenstrom vorhanden ist¹⁵.

Mit *Mixing Level* kann die Wiedergabe weiter optimiert werden, indem der Parameter auf den absoluten SPL-Wert gesetzt wird, der bei der finalen Mischung vom Toningenieur verwendet wurde.

Room Type beschreibt die Art der Mischung bzgl. der verwendeten Equalizer-Kurve. *Large-Room* bezeichnet dabei die „Dubbing Stage“ mit X-Curve Equalizer¹⁶, *Small-Room* ein gewöhnliches Tonstudio mit flachem Frequenzgang.

DC Filter, *Low Pass Filter* und *LFE Lowpassfilter* sorgen dafür, dass bei der Kodierung keine unerwünschten Frequenzen auftauchen.

Surround 3 dB Attenuation senkt die Surroundkanäle vor der Kodierung um 3 dB; dieser Parameter wird bei Audiomaterial gesetzt, das im Kino vorgeführt werden soll.

Bei der Distribution von Digitalfernsehen kann mit Hilfe der *Parameter Audio Service Types* oder auch *Bitstream Mode* das Audiosignal näher beschrieben werden. Gekennzeichnet wird u.a., ob es sich bei dem Audiosignal um einen vollständigen Soundtrack incl. Dialoge und Musik, um Effekte und Musik oder eine alternative Dialogspur handelt. Es ist somit möglich, mehrere Dolby Digital Datenströme zu einem Gesamtsignal zu kombinieren.

(vgl. Dolby [5], Dolby [6])

¹⁵ Es ist nur in bestimmten Channel Modes möglich, einen zusätzlichen LFE-Kanal zu benutzen; z.B. ist ein Mono-Kanal und ein LFE-Kanal nicht möglich.

¹⁶ Da die akustische Dämpfung der Höhen im Kino sehr hoch ist, werden die Höhen im Mix leicht angehoben. Die Höhe der Anhebung wird als X-Curve bezeichnet und unterliegt einem internationalen Standard.

3.2 Professionelle Formate - Dolby E

Dolby E ist ein professionelles Format für Mehrkanalproduktionen, welches bei Produktion, Post-Produktion, Editing und In-House-Distribution verwendet wird und seit 1999 im Einsatz ist.

Ein Audiodatenstrom mit einer Datenrate von 1,92 Mbit/s (2 Kanäle mit 20 Bit/ 48 kHz) kann dabei bis zu acht diskrete Audiokanäle mit Metadaten für jeden einzelnen Audiokanal und SMPTE-Timecode über zwei AES-3 Leitungen übertragen. Üblicherweise wird eine Kombination aus einem 3/2/1-Stereo- und einem 2/0-Stereosignal verwendet. Es sind aber auch andere Kombinationen wie z.B. drei 2/0-Stereosignale oder sechs Monosignale möglich. Dolby E ermöglicht die Nutzung der bisherigen Audioinfrastruktur von Produktionsstudios und kann auf allen gängigen Formaten aufgezeichnet werden.

Im Gegensatz zu Dolby Digital sollen mit Dolby E bis zu zehn Generationen von Kodierungen und Dekodierungen ohne Qualitätsverluste möglich sein¹⁷.

Die Dolby E Audioframes sind an die gängigen Video Frame Raten (30; 29,97; 25; 24; 23,98 bei 48 kHz Abtastrate) angepasst und ermöglichen ein „Frame-genaues“ Schneiden in codierter Form. Die Latenz beträgt dabei einen Frame pro Kodierung und Decodierung; das Bild muss daher durch ein Schwarzbild-Signal angeglichen werden.

Dolby E unterstützt ebenso wie Dolby Digital Metadaten (siehe oben); es ist somit möglich diese Informationen während der gesamten Produktionskette bis hin zum Konsumenten zur Verfügung zu stellen.

Dolby E Daten werden direkt in den Audiodatenteil eines AES-3 Subframes integriert. Jeder Dolby E Frame beginnt mit einem Burst Header oder auch Guard Band gefolgt von einem Synchronisationssignal, Metadaten, dem kodierten Audiomaterial und Zusatzdaten (siehe Abbildung 3-11). Das Synchronisationssignal wird im SPTE 337M und 338M spezifiziert. Jeder Wechsel zwischen unterschiedlichem Audiomaterial muss am Anfang jedes Frames innerhalb des Guard Bandes erfolgen (vgl. Magarelli, S. 3).

Da sich Dolby E Daten zwar über herkömmliche AES-3 Leitungen übertragen lassen, es sich dabei aber nicht um Audiosignale, sondern um Daten handelt, muss der gesamte Signalweg frei von Bearbeitungen wie z.B. Lautstärkeanpassungen, Sample-Raten-Konvertern oder anderen Fehlerquellen sein.

¹⁷ Carsten Adam hat dazu in seiner Diplomarbeit festgestellt, dass bereits ab dem fünften Zyklus wahrnehmbare Artefakte festzustellen sind (vgl. Adam, S. 116).

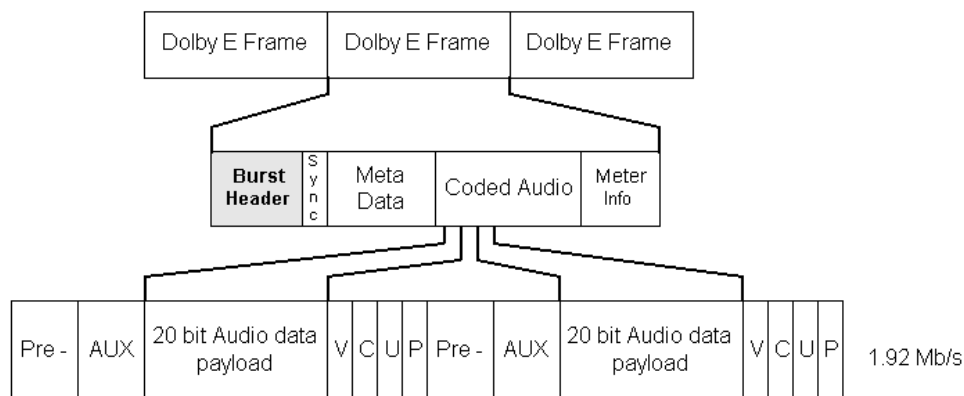


Abbildung 3-11 : AES-3 /Dolby E Datenstruktur (vgl. Magarelli, S.2)

Im deutschsprachigen Raum war das ORF mit dieser Übertragungstechnik Vorreiter. Das Neujahrskonzert Silvester 2002, wurde erfolgreich in dieser Technik übertragen und mit dem Dolby Digital Format vom Sendezentrum per Satellit und Kabel zum Konsumenten transportiert. Im September 2003 übertrug SAT.1 dann ein „Champions League“-Fußballspiel. Das ZDF entschied sich 2004, die aktuelle Staffel der Show „Wetten, dass...?“ in Dolby Digital 3/2/1-Stereo auszustrahlen. Hierzu setzte das ZDF das erste Mal die Dolby E Technik ein, um Mehrkanalaudio zum Sendezentrum zu übertragen (vgl. Adam, S.26).

(vgl. Carroll; Dolby [9]; Krämer, S. 63-67; Magarelli; Pohlmann, S.357)

4 Downmix-Technik

Dieses Kapitel befasst sich mit einer weiteren Variation des automatischen Downmix sowie den Folgen für das Audiomaterial. Hierzu zählen technische Fehler, Auswirkungen auf die Darstellung der Räumlichkeit sowie die Richtungsabbildung und Bewegung bei unterschiedlichen Stereophonieverfahren. Zum Schluss werden die Folgen des Downmix auf die Aufteilung des Klangfeldes und Auswirkungen auf die Dramaturgie diskutiert.

4.1 ITU-R 775 Downmix

Neben dem in Kapitel 3 erläuterten Dolby Downmix ist auch ein Downmix nach ITU-R 775 üblich. Der Downmix wird mit Left Only/ Right Only (Lo/Ro) bezeichnet und ist im Gegensatz zum Lt/Rt-Downmix kein Matrix-Verfahren, das es ermöglicht aus dem 2/0-Stereosignal wieder ein Mehrkanalsignal zu gewinnen.

Der Lo/Ro-Downmix findet Verwendung als Alternative zum Lt/Rt-Downmix auf der DVD-Audio, bei Dolby Digital Signalen oder auch in Computerformaten, wie z.B. dem WMA-Format der Firma Microsoft. Der Downmix lässt sich mittels *Downmix-Koeffizienten* anpassen.

Die allgemein gültige Formel der ITU für den Downmix lautet:

$$L_o = L + 0,7 \cdot C + k \cdot L_s$$

$$R_o = R + 0,7 \cdot C + k \cdot R_s$$

Der Faktor k ist der Downmix-Koeffizient für die Surroundkanäle und wird im allgemeinen 0,7 gewählt (vgl. Theile, 2003, S.19).

Die Formel stellt dabei keine feste Regel dar, sondern der Downmix ist individuell an das Audiomaterial anzupassen. So empfiehlt es sich z.B. bei der Erstellung von DVD-Audio-Medien alle Audiokanäle in der Lautstärke zu verringern, um ein Übersteuern des Downmix zu vermeiden (vgl. The Recording Academy's Producers & Engineers Wing, S. 5-12).

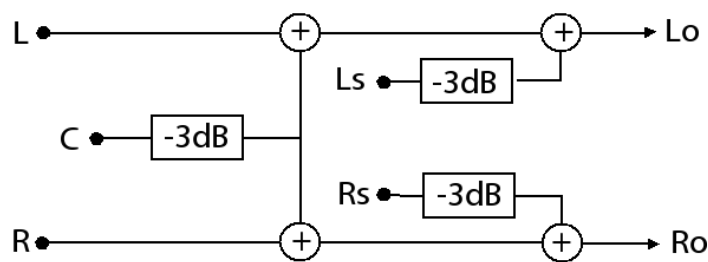


Abbildung 4-1 : ITU-R 775 Downmix mit $k=0,7$

Genau wie beim Lt/Rt-Downmix soll der LFE-Kanal nur zum Downmix hinzugefügt werden, wenn dieser ergänzende Elemente enthält, die nicht in den übrigen Audiokanälen vorhanden sind.

Eine Besonderheit des ITU-Downmix

entsteht durch die Möglichkeit bei Kodierung in das Dolby Digital Format die Surroundkanäle in der Phase um 90° dauerhaft zu drehen, um dem Decoder die Möglichkeit zu bieten, einen Pro Logic kompatiblen Downmix zu erstellen¹⁸. Da der

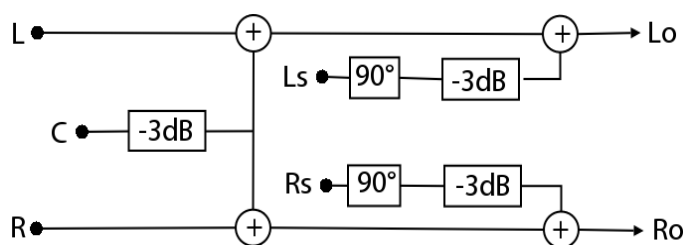


Abbildung 4-2 : ITU-Downmix mit $k=0,7$ und Dolby® Digital Metadata Surround Phase Shift = enabled

Decoder nicht explizit bei gedrehter Phase der Surroundkanäle einen Downmix nach Dolby Pro Logic erstellen muss, ergibt sich ein Downmix nach dem Schema in Abbildung 4-2.

¹⁸ siehe Kapitel 3.1.2.3 Metadaten

Durch die Phasendrehung gelten für den Downmix die Gesetzmäßigkeiten der Addition von Audiosignalen mit einer Phasenverschiebung von 90° ¹⁹.

Bei der diskreten 3/2-Stereo-Wiedergabe sorgt die Phasendrehung jedoch für eine Dekorrelation der Surroundkanäle zu den vorderen Audiokanälen und damit für eine Destabilisierung der seitlichen Phantomschallquellen; das Klangbild scheint insgesamt „breiter“ zu werden. Anhand von Testsignalen (Musik, rosa Rauschen, Sinussignale unterschiedlicher Frequenz) kann man sich leicht davon überzeugen, dass die Verbreiterung des Klangbildes frequenzabhängig und in einigen Fällen hörbar ist. Die Frequenzabhängigkeit lässt sich einfach erklären. Es gilt: $\rho = 360^\circ \cdot f \cdot \Delta t = 90^\circ$ daraus folgt

$\Delta t = \frac{1}{4f}$. Je geringer also die Frequenz, desto größer die Zeitverschiebung durch die

Phasenverschiebung und damit nach Kapitel 2.5 auch die Verschiebung der Phantomschallquelle für die jeweilige Frequenz. Zur Zeit der Entwicklung der Dolby Digital Kodierung wird dieser Tatsache vermutlich von den Entwicklern keine große Relevanz zugesprochen worden sein, da die Surroundkanäle vornehmlich zur Erzeugung von Diffusschall gedacht waren. Für das Verhältnis der Surroundkanäle untereinander hat die Phasenverschiebung keine Auswirkung, da beide Kanäle gleich verzögert werden. Im weiteren soll dieser Downmix mit „ITU-90“ bezeichnet werden.

4.2 Weitere Möglichkeiten des Downmix

Eine weitere Möglichkeit des Downmix bietet u.a. das „Spatial Audio Coding“, wie es für das neue Tonformat *mp3-Surround* spezifiziert wurde. Da allerdings zum momentanen Zeitpunkt unklar ist, inwieweit sich dieses Format zur Distribution außerhalb des Internets oder z.B. in der Anwendung in Downloadplattformen, durchsetzen wird, soll hier nicht weiter darauf eingegangen werden²⁰.

Zusätzlich zu den hier vorgestellten Methoden eines Downmix, der darauf angelegt ist eine Stereomischung zu erstellen, die einer „handgemachten“ Version möglichst nahe kommt, werden von verschiedenen Firmen Lösungen angeboten, die versuchen mit einer Wiedergabe über zwei Lautsprecher oder auch Kopfhörer einen Klangeindruck zu schaffen, der der Wahrnehmung einer Präsentation im 3/2-Stereo-Format entspricht. Dazu werden

¹⁹ siehe Kapitel 2.4

²⁰ Weitere Informationen unter: <http://www.ils.fraunhofer.de/amm/download/mp3surround/>

z.B. die Reflektionen an den Wänden oder auch die Außenohrübertragungsfunktion nachgebildet.

Die bekanntesten Formate stammen von der Firma Dolby und werden unter den Namen Dolby Virtual Speaker²¹ und Dolby Headphone²² vertrieben. Das neueste Produkt dieser Art wurde vom Fraunhofer Institut für Integrierte Schaltungen entwickelt und nennt sich Ensonido[®] ²³.

Diese Downmix-Techniken seien hier nur der Vollständigkeit halber erwähnt, jedoch nicht im Detail erläutert.

4.3 Probleme beim Downmix

Der Downmix kann unter bestimmten Voraussetzungen zu Problemen führen. Neben technischen Fehlern ergeben sich Nachteile durch die Reduktion der Audiokanäle im Bezug auf die Abbildung von Räumlichkeit, die Dramaturgie und die ursprüngliche Aufteilung des Klangfeldes.

Beim Dolby Pro Logic II Downmix ist im Gegensatz zum ITU-Downmix keine eindeutige „Kanaltrennung“²⁴ der Surroundkanäle nach dem Downmix mehr vorhanden. Eine Erklärung liefert die in Kapitel 3.1.1.1 abgebildete Dolby Pro Logic II Encoder Matrix. Aus dieser geht hervor, dass sowohl der linke, wie auch der rechte Surroundkanal beim Downmix dem Lt- und dem Rt-Kanal addiert werden. Somit befinden sich in *beiden* Audiokanälen der Lt/Rt-Mischung Signalanteile sowohl des linken als auch rechten Surroundkanals.

Der Dolby Pro Logic Downmix kann durch die Addition der beiden Surroundkanäle zu einem Monosignal zu Verlusten der Klangqualität führen.

Durch einen Downmix entfallen mindestens drei Kanäle des 3/2-Stereobildes und die darin enthaltenen Informationen werden den Frontkanälen zugemischt. Um die Folgen dieser Reduktion einzuschätzen, ist es wichtig sich der elementaren Funktionen des Center- und der Surroundkanäle bewusst zu sein. Der Centerkanal soll vor allem für eine stabile Richtungswahrnehmung in den vorderen Lautsprechern sorgen. Er dient dabei als „Ankerpunkt“ für Schlüsselkomponenten des Gesamtklangbildes (vgl. [P&E], S.4-5).

²¹ Weitere Informationen unter: http://www.Dolby.com/consumer/technology/virtual_speaker.html

²² Weitere Informationen unter: <http://www.Dolby.com/consumer/technology/headphone.html>

²³ Weitere Informationen unter: <http://www.iis.fraunhofer.de/amm/techinf/ensonido/index.html>

²⁴ Der Begriff „Kanaltrennung“ im Zusammenhang mit dem Dolby Pro Logic II Downmix soll in dieser Arbeit den oben erläuterten Zusammenhang bezeichnen.

Die Surroundkanäle haben im wesentlichen folgende Hauptaufgaben zu erfüllen:

Die Darstellung von Räumlichkeit, Atmosphäre und Effekte sowie die Ergänzung und Unterstützung der vorderen Audiokanäle (vgl. ebd., S.4-6).

Die nun folgenden Punkte sollen erläutern, welche Auswirkungen der Downmix auf die Informationen von Center und Surroundkanälen haben kann.

4.3.1 Technische Fehler

Der Downmix führt unter bestimmten Voraussetzungen zu unterschiedlichen technischen Fehlern.

Einer dieser Fehler entsteht aus der Multichannel-Panpot-Matrix bei Intensitätsstereophonie. Ein Monosignal, das sich in der Mittelstellung zwischen zwei Kanälen befindet, erfährt eine Dämpfung von 3 dB. Die Multichannel-Panpot-Matrix entspricht dabei dem Schema in Abbildung 4-3. Bei einem Downmix nach ITU ergibt sich nun für Signale, die aus mehreren Kanälen gleichzeitig wiedergeben werden, ein Fehler, der zu Übersteuerungen führen kann.

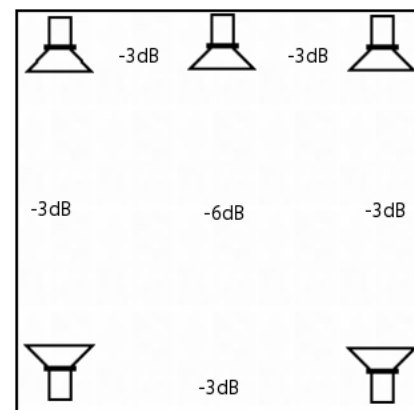


Abbildung 4-3 : Multichannel-Panpot-Matrix (vgl. Rathgeber, S. 42)

Als Beispiel sei ein Signal von 0 dB angeführt, dessen Phantomschallquelle sich in der Mitte zwischen linkem Audiokanal und Center befinden soll. Hiefür gilt ²⁵:

$$L = -3 \text{ dB} \approx 0,7 \text{ V}; C = -3 \text{ dB} \approx 0,7 \text{ V}$$

Daraus folgt mit einem $k=0,7$:

$$L_o = L + 07 \cdot C + k \cdot LS = 0,7 \text{ V} + 0,7 \cdot 0,7 \text{ V} = 1,2 \text{ V} \approx \underline{+1,64 \text{ dB}}$$

$$R_o = R + 0,7 \cdot C + k \cdot RS = 0,7 \cdot 0,7 \text{ V} = 0,5 \text{ V} = -6 \text{ dB}$$

Insgesamt ergibt sich für das Ohr ein Leistungssummenpegel von:

$$10 \text{ dB} \cdot \log (L_o^2 + R_o^2) = 10 \text{ dB} \cdot \log (1,2^2 + 0,5^2) = \underline{+2,3 \text{ dB}}$$

Für den Downmix eines Signals, das zwischen dem linken Audiokanal und dem Center hin- und hergepannt wird, bedeutet dies, dass die Lautstärke im Lo-Audiokanal bis zur Mitte ansteigt und ab dann wieder abfällt. Abhilfe würde eine veränderte Downmix-Matrix schaffen, wie sie in einigen Surround Applikationen eingestellt werden kann; mit einer

²⁵ mit $L = 20 \text{ dB} \cdot \log \left(\frac{U}{U_o} \right)$ mit $U_o = 1 \text{ V}$

Dämpfung von 6 dB würde ein Übersteuern verhindert werden, allerdings bedeutet dieses für den Mehrkanal-Mix, dass ein Signal an den entsprechenden Panorama-Positionen eine Dämpfung von 3 dB erfahren würde (vgl. Wittek, Folie 9; Rathgeber, S. 41-45). Dieser Fehler tritt bei einem Signal zwischen Center und linkem/ rechtem Audiokanal sowohl beim ITU-, als auch beim Dolby Pro Logic/ Pro Logic II Downmix auf.

Für Signale zwischen linkem Surroundkanal und linkem Audiokanal bzw. rechtem Surroundkanal und rechtem Audiokanal ergibt sich beim ITU-Downmix mit einem $k=0,7$ ein Fehler von 1,5 dB. Beim Dolby Pro Logic II und ITU-90° Downmix wird durch die Phasenverschiebung von 90° das Pegelproblem behoben, da die Energie beim Downmix zwischen dem linken Surroundkanal und dem linken Audiokanal, bzw. dem rechten Surroundkanal und dem rechten Audiokanal konstant bleibt (siehe Kapitel 2.4) (vgl. Wittek, Folie 12).

Ein ähnliches Problem wie der Pegelfehler bei Signalen in der Phantommitte betrifft ITU- und Dolby Downmix gleichermaßen (siehe Abbildung 4-4):

Bei einem Audiosignal mit einem Pegel von beispielsweise 0 dB im linken Audiokanal und Center ergibt sich:

$$L = 1V = 0 \text{ dB} \quad \text{und} \quad C = 1V = 0 \text{ dB.}$$

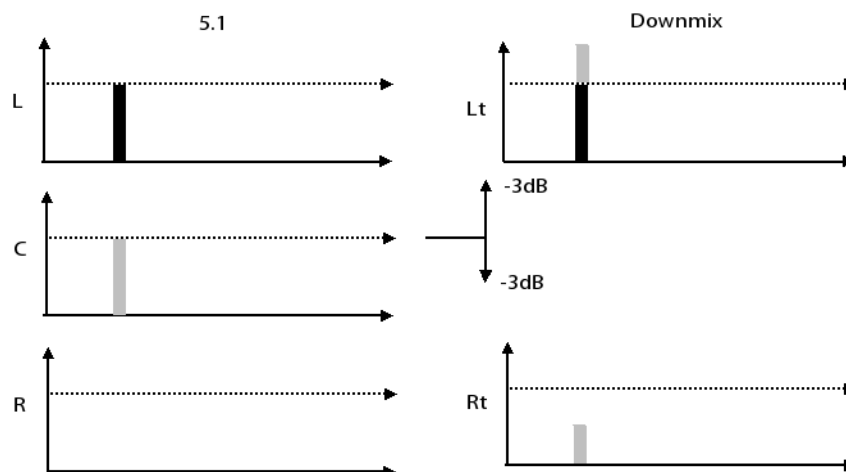


Abbildung 4-4 : Übersteuern durch Downmix (vgl. Wittek, Folie 8)

Durch den Downmix wird das Signal im Center nach einer Dämpfung um 3 dB dem rechten und linken Audiokanal gleichermaßen addiert, so dass sich für das Lt-Signal ergibt:

$$L_t = L + 0,7 \cdot C = 1V + 0,7 \cdot 1V = +1,7V \quad (\text{vgl. Wittek, Folie 8})$$

Somit kann es durch den Downmix zu Übersteuern kommen.

4.3.2 Abbildung von Räumlichkeit bei der Wiedergabe

Die Wahrnehmung von Räumlichkeit und räumlicher Tiefe beruht im wesentlichen auf dem Verhältnis der Verzögerung und Lautstärke der frühen und späten Reflektionen sowie des Nachhalls zum Direktschall. Nach Blauert erzeugt der Primärschall ein Primärhörereignis und löst nach dem Gesetz der ersten Wellenfront einen Inhibitionseffekt aus, welcher für eine gewisse Zeit weitere Hörereignisse unterdrückt. Nach Ablauf dieser Zeitspanne erzeugen die durch den Nachhall hervorgerufenen, weitgehend inkohärenten Ohrsignale ein diffus lokalisiertes Hörereignis (Nachhallhörereignis), das je nach Nachhallzeit des Raumes langsam abklingt (vgl. Blauert, 1974, S. 233). Da der direkte Schall bei unbekanntem Schallquellen bis auf eine Absorption von Frequenzanteilen ab etwa 10 kHz (vgl. ebd., S. 102) nur wenige Informationen über die Entfernung der Schallquelle enthält, sind vornehmlich die indirekten Anteile des Gesamtschalls für die Darstellung von räumlicher Tiefe verantwortlich²⁶. Die Differenz der Pegel von Primärschall und reflektiertem Schall liefert dem Gehör Informationen über die Schallquellenentfernung, die bei der Bestimmung der Hörereignisrichtung berücksichtigt werden. Je mehr man sich in einem geschlossenen Raum einer Schallquelle nähert, um so stärker wird der Anteil des Primärschalls gegenüber demjenigen des diffusen Feldes (vgl. ebd., S. 223).

Räumlichkeit bedeutet weiter, dass eine charakteristische Aufweitung der Hörereignisse geschieht, so dass diese einen größeren Raumbereich ausfüllen, als durch die visuellen Konturen der Schallquelle gegeben ist (vgl. ebd. (1985), S. 75). Voraussetzung für das Auftreten von Räumlichkeit ist, dass die Ohrsignale unkorreliert sind (vgl. ebd. (1985), S. 76; (1974), S. 194). Auch künstlicher Nachhall klingt natürlicher, wenn sich dieser aus unkorrelierten Signalen zusammensetzt (vgl. Griesinger: A Note to Sengpiel).

Experimentelle Versuche haben ergeben, dass der Grad der räumlichen Empfindung vom Einfallswinkel der seitlichen Reflektionen abhängt, (vgl. Blauert, 1985, S. 82) und dass allein durch in der horizontalen Ebene aufgestellte Lautsprecher ein natürlicher Raumeindruck geschaffen werden kann (vgl. Theile, 2001, 1. Version, S. 5). Diese Umstände können bei der Produktion von Programmen im 3/2-Stereo-Format genutzt werden, um durch die Reproduktion von Reflektionen und Nachhall in den Surround-Lautsprechern einen weitgehend natürlichen Raumeindruck zu schaffen.

²⁶ Beim natürlichen Hören in normalen Alltagssituationen spielt selbstverständlich auch die Intensität des Schalls eine Rolle für die wahrgenommene Entfernung; allerdings nur, wenn zugleich eine visuelle Entsprechung der Schallquelle vorliegt. Zudem ist die wahrgenommene Lautstärke bei frontalem Schalleinfall frequenzabhängig, wie aus den Kurven gleicher Lautstärke hervorgeht.

Ein dreidimensionaler Raumeindruck benötigt also einen Schalleinfall an der Seite des Hörers. Durch verschiedene Untersuchungen²⁷ ist inzwischen nachgewiesen worden, dass seitliche Phantomschallquellen sehr instabil sind und stark vom Spektrum des Signals, sowie der Position des Hörers abhängen. Zudem weisen sie im Gegensatz zu realen Schallquellen Klangverfärbungen auf. Ein natürlicher Raumeindruck ist mit einer 3/2-Stereo-Standard-Aufstellung jedoch ohne größere Schwierigkeiten realisierbar, da stabile seitliche Phantomschallquellen für die Erzeugung von Räumlichkeit nicht vorausgesetzt werden müssen und störende Klangverfärbungen im Diffusfeld untergehen (vgl. Theile, 2001, S. 25).

Beim natürlichen Hören und der 3/2-Stereo-Wiedergabe trifft der Nachschall im Gegensatz zum Direktschall aus verschiedenen Richtungen ein und erzeugt so einen umhüllenden Raumeindruck. Bei der Wiedergabe über die 2/0-Stereo-Standard-Aufstellung ist die Distanz zu eventuellen Phantomschallquellen gleich der Entfernung des Hörers zu den beiden Lautsprechern. Die räumliche Perspektive kann dabei nur als Simulation zwischen den beiden Lautsprechern hergestellt werden, ähnlich der zeichnerischen Darstellung von drei-dimensionalen Objekten (vgl. Theile, 1991, S.2). Um eine räumliche Illusion zu schaffen, müssen daher in Analogie zu den oben erwähnten Mechanismen der Wahrnehmung von Räumlichkeit die frühen Reflektionen und der Nachhall der Schallquelle in der 2/0-Stereo-Mischung enthalten sein. Entweder werden diese schon in der Aufnahme erzeugt (z.B. durch Raummikrophone) oder sie müssen mit Hilfe von digitalen Signalprozessoren, mechanisch arbeitenden Hallgeräten oder speziellen Hallräumen simuliert werden. Enthält ein Klang mehr indirekte als direkte Signalanteile, so nimmt das Ohr ein Hörereignis wahr, das hinter der Lautsprecherachse zu liegen scheint.

Beim Downmix erfolgt nun ein Wechsel von einer natürlichen Wahrnehmung des Raumes zu einer simulierten Räumlichkeit (vgl. Theile, 2001, S.22). Theile hat diesen Unterschied in der Abbildung 4-5 dargestellt.

²⁷ siehe z.B.: Ziegelmeier und Theile: Darstellung seitlicher Schallquellen bei Anwendung des 3/2-Stereo-Formates.

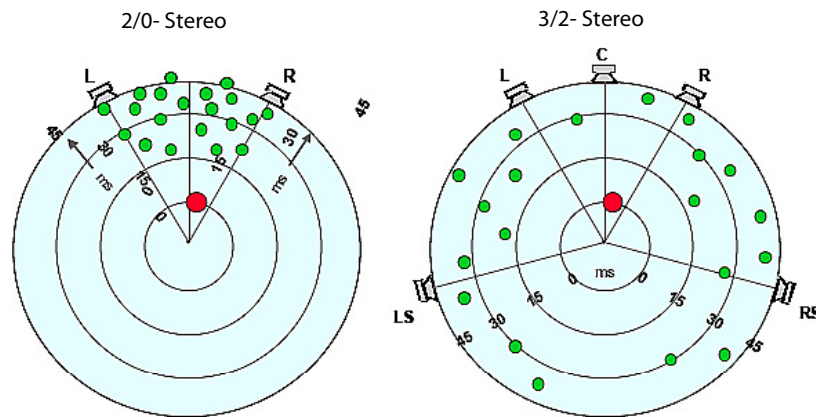


Abbildung 4-5 : Simulation von Räumlichkeit²⁸. (Theile, 2001, S. 22).

Durch die Addition der Surround-Informationen zu den vorderen Audiokanälen bleiben die Rauminformationen vollständig erhalten; der Downmix kann jedoch je nach Art der Zusammensetzung der Signale störend wirken. Es besteht zum Beispiel die Möglichkeit, dass das Gesamtsignal zu sehr verhallt wirkt. Oder andersherum: Das Signal kann durch die Addition „trockener“ wirken, da sich die phasengleichen Anteile des Direktschalls addieren und die Raumsignale mit dekorrelierten Phasen im Verhältnis leiser wirken. Daher sollte beim Downmix darauf geachtet werden, mit welchem Pegel die Surroundkanäle addiert werden.

Die Tatsache, dass beim Dolby Pro Logic II Downmix keine eindeutige Kanaltrennung der Surroundkanäle nach dem Downmix mehr vorhanden ist, wird für die Wahrnehmung der Räumlichkeit nach dem Downmix keine besondere Bedeutung haben, da es sich vorwiegend um unkorrelierte Signale handelt.

Beim Downmix nach Dolby Pro Logic muss beachtet werden, dass die beiden Surroundkanäle zu einem Monosignal addiert werden und sich somit der gesamte Raumanteil der Surroundkanäle nun in der Stereomitte befindet. Das 2/0-Stereobild wirkt dadurch wesentlich „trockener“.

Bei der diskreten 3/2-Stereo-Wiedergabe eines Dolby Pro Logic kompatiblen Dolby Digital Signals kann unter Umständen der Eindruck entstehen, dass das Klangbild im Gegensatz zu einem Signal mit ursprünglicher Phasenlage indirekter wird. Dies liegt daran, dass die Dekorrelation durch die 90° Phasenverschiebung der Surroundsignale, sich vornehmlich hörbar auf die direkten Anteile des Gesamtklangs auswirkt, wohingegen sich der Diffusschall in der Regel aus dekorrelierten Signalen zusammensetzt.

²⁸ Die Zeitskala in ms definiert die Ankunftszeit des Klangs beim Hörer. Die Punkte stellen das Primärschallereignis (Rot) und Reflektionen (Grün) dar.

4.3.3 Richtungsabbildung und Bewegung

Nach Kapitel 2.5 werden, um die Bewegung eines Hörereignis bei der Wiedergabe über Lautsprecher zu erzeugen, mindestens zwei reale Schallquellen benötigt.

Verändert man innerhalb bestimmter Grenzen kontinuierlich die Pegel oder/ und Laufzeitdifferenzen der Signale, so wandert die Phantomschallquelle entlang der Lautsprecherbasis bis sie schließlich in einem der beiden Lautsprecher stehen bleibt (vgl. Dickreiter, Band1, S. 124).

Hier sollen nun die Auswirkungen des Downmix auf Bewegungen von Phantomschallquellen innerhalb des 3/2-Stereobildes bei Intensitäts-, Laufzeit- und Äquivalenzstereophonie diskutiert werden.

4.3.3.1 Richtungsabbildung und Bewegung mit Hilfe von Pegeldifferenzen

Richtungsabbildung mit Hilfe von Pegeldifferenzen findet sich in der Audiotechnik in so gut wie jedem Mischpult oder jeder digitalen Audioworkstation (DAW) als eine Grundlage der Klanggestaltung wieder. In der Mikrofon-Aufnahmetechnik beruhen Verfahren nach dem Prinzip der Intensitätsstereophonie auf der Ausnutzung von Pegeldifferenzen.

Strahlen bei der Wiedergabe zwei Lautsprecher dasselbe Signal phasengleich ab, so wird bei gleichem Pegel die Phantomschallquelle genau in der Mitte der Lautsprecherbasis wahrgenommen. Bei Pegeldifferenzen wandert die Phantomschallquelle auf der Basis seitlich aus, um schließlich ganz am Ort des Lautsprechers mit dem höheren Pegel stehen zu bleiben (vgl. Dickreiter, Band 1, S. 127). Die Tabelle 4-1 gibt einen Überblick über den Zusammenhang von Auslenkung des Hörereignisses und der Pegeldifferenz bei einer 2/0-Stereo-Standard-Aufstellung.

| | | | | | |
|---------------------------|------|------|--------|-------|-------|
| Hörereignisrichtung | 0 % | 25 % | 50 % | 75 % | 100 % |
| Pegeldifferenz ΔL | 0 dB | 3 dB | 6,5 dB | 11 dB | 18 dB |

Tabelle 4-1: Zusammenhang von Auslenkung des Hörereignisses und der Pegeldifferenz²⁹

Um nun eine Phantomschallquelle von einem Lautsprecher zum anderen gradlinig wandern zu lassen, muss die Summe der beiden akustischen Leistungen stets konstant

²⁹ Eine Auslenkung von 0 % entspricht dabei der Mitte der Wegstrecke zwischen zwei Lautsprechern sowie 100 % der Lokalisation in nur einem der beiden Lautsprecher. Die anderen Werte entsprechen einer linearen Einteilung der Wegstrecke zwischen 0 % und 100 %. (vgl. Sengpiel, Vergleich von Pegeldifferenz-Lokalisationskurven bei Interchannelsignalen)

bleiben, d.h. die Gesamtlautstärke darf sich während des Panning-Vorgangs nicht ändern (vgl. Sengpiel, Panoramaregler).

Bewegungen von Phantomschallquellen durch Intensitätsunterschiede sind in der Postproduktion leicht zu realisieren. Alle dem Autor bekannten „surroundfähigen“ DAWs sind mit Multichannel-Panpots ausgestattet, die auf der in Abschnitt 4.3.1 vorstellten Multichannel-Matrix beruhen. Mit Hilfe der Automation können auf diese Weise leicht Bewegungen von Phantomschallquellen realisiert werden. Die nun folgenden Anmerkungen zur Intensitätsstereophonie gelten meiner Ansicht nach daher vornehmlich für Produktionen, die ihre eigentliche Klanggestaltung in der Postproduktion erfahren haben und weniger für Produktionen, die eine Dokumentation eines live dargebotenen Klangerlebnisses darstellen.

Beim Panning in der 3/2-Stereo-Anordnung mit Hilfe von Pegeldifferenzen treten nach dem Downmix zunächst die in Abschnitt 4.3.1 angesprochenen technischen Fehler auf. Signale die zwischen Center und dem linken oder rechten Lautsprecher und solche, die zwischen den vorderen und hinteren Lautsprechern hin- und herwandern, können nach dem Downmix im Pegel an- und wieder absteigen.

Der Downmix der vorderen Kanäle führt zunächst zu einer Monosummenbildung durch die Addition des Centerkanals zum linken und rechten Audiokanal. Es gelten somit die Bedingungen für Monokompatibilität, die bei reiner Intensitätsstereophonie in der Regel gegeben sind.

Die Abbildung der vorderen Audiokanäle erscheint nach dem Downmix breiter (vgl. Wittek, Folie 11). Durch die Dämpfung des Centerkanals beim Downmix wird bei einem Signal zwischen Center und Links, bzw. Center und Rechts der linke bzw. rechte Audiokanal im Verhältnis zum Original im Pegel größer, so dass das Hörereignis weiter links bzw. rechts lokalisiert wird. Die Bewegungsrichtungen bleiben erhalten.

Obwohl, wie bereits erwähnt wurde, seitliche Phantomschallquellen sehr instabil sind, ist es dennoch möglich bei 3/2-Stereo zwischen den hinteren und vorderen Lautsprechern seitliche Bewegungen von Phantomschallquellen zu erzeugen. Ich konnte dieses anhand eines Hörbeispiels gut nachvollziehen. Mit großer Wahrscheinlichkeit hängt der Erfolg seitlicher Bewegungen von Phantomschallquellen dabei von der Art des Signals ab. Ähnliche Erfahrungen schildert Blauert zum Beispiel im Zusammenhang mit dem Richtungshören in der Medianebene (vgl. Blauert, 1974, S. 85). Bekannte Signale wie z.B. U-Bahn-Geräusche dürften vom Hörer leichter als seitliche, zusammenhängende

Hörereignisse empfunden werden, als unbekannte. Auch die Dauer und der Frequenzumfang eines solchen Hörereignisses werden für die Lokalisation eine wichtige Rolle spielen. Je breiter das Spektrum einer Schallquelle ist, desto mehr Informationen kann das Gehör bezüglich der Schallquellenlokalisierung entnehmen (vgl. ebd., S. 83). Ein Hörereignis, das aus einem lang andauernden, stetigen Geräusch besteht, wird vom Hörer leichter als homogen empfunden werden als eine impulshafte Schallquelle, von der der Hörer eher den Eindruck bekommen kann, dass das Hörereignis zwischen den Lautsprechern hin- und herspringt.

Beim Downmix wird zunächst aus einem „seitlichen 2/0-Stereosignal“ ein Monosignal, da die Surroundsignale zu den vorderen Audiosignalen addiert werden. Es gelten daher die Regeln zur Monokompatibilität.

Durch die Phasenverschiebung von 90° des Surroundkanals beim Dolby Pro Logic II und ITU- 90° -Downmix wird der technische Fehler der Lautstärkezunahme verhindert. Es ergeben sich die in Kapitel 2.4 angesprochenen Vor- und Nachteile der Phasenverschiebung von 90° : Die Gesamtleistung bleibt gleich und die Auslöschung von Signalen, die eine Phasendifferenz von 180° aufweisen, kann vermieden werden. Es kann aber auch zu neuen Auslöschungen kommen, wenn Signale erst durch die Phasenverschiebung eine Phasendifferenz von 180° erhalten.

Durch die Dekorrelation der beiden Signale ergeben sich für verschiedene Frequenzen nach Punkt 4.1 unterschiedliche Zeitverschiebungen. Das bedeutet, dass im Falle eines Dolby Pro Logic II und ITU- 90° Downmix bei Intensitätsstereophonie in einem gewissen Rahmen auch die Auswirkungen des Downmix bei Laufzeitstereophonie gelten³⁰.

Die seitlichen Bewegungen von Phantomschallquellen gehen durch den Downmix verloren. Sie können aber unter Umständen in anderer Form auftreten.

Bei der 2/0-Stereo-Wiedergabe sind Phantomschallquellen vor der Lautsprecherachse nicht möglich, also auch keine Hörereignisse seitlich zum Hörer. Bewegungen können sich daher nur zwischen den beiden Lautsprechern oder hinter der Lautsprecherbasis abspielen. Bewegt sich bei einer Stereoaufnahme die Schallquelle von den Mikrofonen weg, so kann diese Art der Bewegung auch während der Wiedergabe vom Hörer nachvollzogen werden. Einen ähnlichen Effekt erhält man, wenn man in der Postproduktion das Verhältnis von Diffusschall und Originalsignal kontinuierlich verändert.

³⁰ siehe Kapitel 4.3.3.2

Enthält die 3/2-Stereo-Produktion genügend indirekte Anteile, so kann unter Umständen der Downmix nach ITU/ ITU-90° oder Dolby Pro Logic II die vormals seitliche Bewegung der Phantomschallquelle zumindest als hin- oder wegbewegen einer Schallquelle darstellen. Ähnlich wie beim Wechsel von einer natürlichen Wahrnehmung des Raumes zu einer simulierten Räumlichkeit, erfolgt hier ein Wechsel von einem „realen“ Eindruck der Bewegung zur Simulation.

Die nicht eindeutige Kanaltrennung der Surroundkanäle nach dem Downmix beim Dolby Pro Logic II Downmix kann dazu führen, dass die Audiosignale im Lt/Rt-Signal nicht mehr stabil nur einer Seite zuzuordnen sind. Da die ursprünglichen Signale stark korreliert waren, können sich im ungünstigsten Fall neue Phantomschallquellen bilden. Das Audiosignal kann daher in der gesamten Stereobreite „verschmiert“ wirken.

Beim Downmix nach Dolby Pro Logic werden die Surroundkanäle zu einem Monosignal addiert, welches sich nun in der Stereomitte befindet. Es gelten damit die in Kapitel 4.3.2. angesprochenen Nachteile bezüglich der Darstellung von Räumlichkeit. Die Phantomschallquelle bewegt sich außerdem nun von einem der beiden Lautsprecher zur Mitte oder umgekehrt.

Da nur wenige Hauptmikrophone für das 3/2-Stereo-Format nach dem Prinzip der Intensitätsstereofonie arbeiten und diesem Verfahren im Allgemeinen eine weniger gute Räumlichkeit zugeschrieben wird, eignen sich diese Aufnahmen wahrscheinlich weniger gut für einen Downmix, wenn seitliche Bewegungen von Phantomschallquellen erhalten bleiben sollen. Gleiches gilt für Bewegungen von Phantomschallquellen, die in der Postproduktion mit Hilfe von Multichannel-Panning und Automation erstellt wurden, wenn das Hörereignis außer Pegeldifferenzen keine weiteren Merkmale über die räumlichen Gegebenheiten enthält.

Bewegungen von Phantomschallquellen zwischen den Surround-Lautsprechern sind mit gewissen Einschränkungen bei 3/2-Stereofonie prinzipiell möglich. Der Winkel zwischen den Surround-Lautsprechern beträgt nach der ITU-775 Empfehlung zwischen 120° und 160° und ist damit im Gegensatz zur Lautsprecher-Aufstellung im 2/0-Stereo-Format deutlich breiter. Phantomschallquellen zwischen den beiden Surround-Lautsprechern sind somit wesentlich instabiler, da sich mit Vergrößerung der Basisbreite die Hörfläche verringert (vgl. Dickreiter, Band 1, S.127). Dennoch sind Phantomschallquellen in den Surround-Lautsprechern möglich (vgl. Dettwiler, S.34). Die Lokalisation von Schallquellen, die sich auf der hinteren Seite des Kopfes befinden, erfolgt dabei durch den Einfluss der

Pinna (vgl. Blauert, 1974, S. 83) und die Auswertung der „richtungsbestimmenden Bänder“ (vgl. ebd., S.89). Zur Unterstützung der „Hinten-Lokalisation“ kann es helfen, die entsprechenden Frequenzbereiche anzuheben (vgl. Sengpiel, Vorteilhafte Filterung der Raumsignale).

Bewegungen von Phantomschallquellen zwischen den Surround-Lautsprechern werden nach dem Downmix nach Dolby Pro Logic II oder ITU/ ITU-90° zu Bewegungen von Phantomschallquellen entlang der Lautsprecherbasis bei 2/0-Stereo-Standard-Aufstellung, wobei wie oben erwähnt die Lautsprecherbasis enger zusammenliegt. Beim Dolby Pro Logic II Downmix kann es durch die schlechte Kanaltrennung der Surround-Informationen nach dem Downmix dazu kommen, dass die Phantomschallquellen nicht mehr stabil sind und breiter wirken.

Beim Downmix nach Dolby Pro Logic wird aus den Surroundkanälen ein Monosignal, somit gelten die Regeln zu Monokompatibilität, sowie die Tatsache, dass Bewegungen entweder verloren gehen oder bei Vorhandensein von genügend Rauminformationen zum Hörereignis einer hin- oder wegbewegenden Schallquelle führen.

4.3.3.2 Richtungsabbildung und Bewegung mit Hilfe von Laufzeitdifferenzen

Richtungsabbildung mit Hilfe von Laufzeitdifferenzen findet in der Praxis eher Anwendung in der Mikrofonaufnahmetechnik als in der Postproduktion. Mir ist lediglich das zur 103. AES Convention in New York von Studer vorgestellte „Virtual Surround Panning“ bekannt³¹. Bei dem verwendeten System werden die Phantomschallquellen wahlweise mit den Signaldifferenzen verschiedener Stereomikrophone, z.B. Laufzeit (A/B), ORTF oder Kugelfläche, positioniert.

In der Mikrofonaufnahmepraxis hingegen sind gerade in der Mehrkanalaudioproduktion Verfahren nach dem Prinzip der Laufzeitstereophonie, vor allem auf Grund ihrer hervorragenden Eigenschaften bezüglich der Abbildung von Räumlichkeit, weit verbreitet. Die Laufzeitstereophonie macht sich den Präzedenz-Effekt zu nutze, indem z.B. bei der Wiedergabe über zwei Lautsprecher ein Kanal gegenüber dem anderen verzögert wird (vgl. Blauert, 1974, S.163). Bei einer Verzögerungszeit von $630\mu\text{s} < t_{\text{Verzögerung}} < 1 \text{ ms}$ hat das Hörereignis die gleiche Richtung wie der früher abstrahlende Lautsprecher erreicht. Erhöht man die Verzögerungszeit über 1 ms hinaus, bleibt die Hörereignisrichtung konstant bis

³¹ Siehe hierzu www.studer.ch → Digitalmischpult D950 M2

zur Echoschwelle, deren Grenzwert von unterschiedlichen Faktoren abhängig ist³² (vgl. ebd. , S.178). Die Tabelle 4-2 gibt einen Überblick über den Zusammenhang von Auslenkung des Hörereignisses und der Laufzeitdifferenz bei einer 2/0-Stereo-Standard-Aufstellung.

| Hörereignisrichtung | 0 % | 25 % | 50 % | 75 % | 100 % |
|------------------------------|------|---------|---------|---------|--------|
| Laufzeitdifferenz Δt | 0 ms | 0,23 ms | 0,48 ms | 0,81 ms | 1,5 ms |

Tabelle 4-2 : Zusammenhang von Auslenkung des Hörereignisses und der Laufzeitdifferenz³³.

Diese Werte wurden empirisch mit breitbandigen Sprach- und Musiksignalen ermittelt und können daher nur als Mittelwerte angenommen werden. Die für eine Auslenkung erforderliche Zeit hängt z.B. auch von der Impulshaftigkeit der Signale ab. So wurde bei Versuchen mit speziellen Stoßsignalen eine Auslenkung von 100 % schon bei einer Laufzeitdifferenz von 0,8 ms nachgewiesen (vgl. Sengpiel, Theoriegrundlagen: Laufzeitstereophonie).

Laufzeiten zwischen zwei Mikrofonen werden im einfachsten Fall durch eine räumliche Trennung der Mikrophone erreicht. In Abhängigkeit von der Mikrofonbasis und des Schalleinfallswinkels entstehen dabei unterschiedliche Laufzeiten.

Bei der Laufzeitstereophonie sind Abbildungs-Nichtlinearitäten zu beachten. So lässt sich bei gleichförmig aufgestellten Schallquellen beobachten, dass die Hörereignisse bei der Wiedergabe in Richtung der Lautsprecher drängen (siehe Abbildung 4-6).

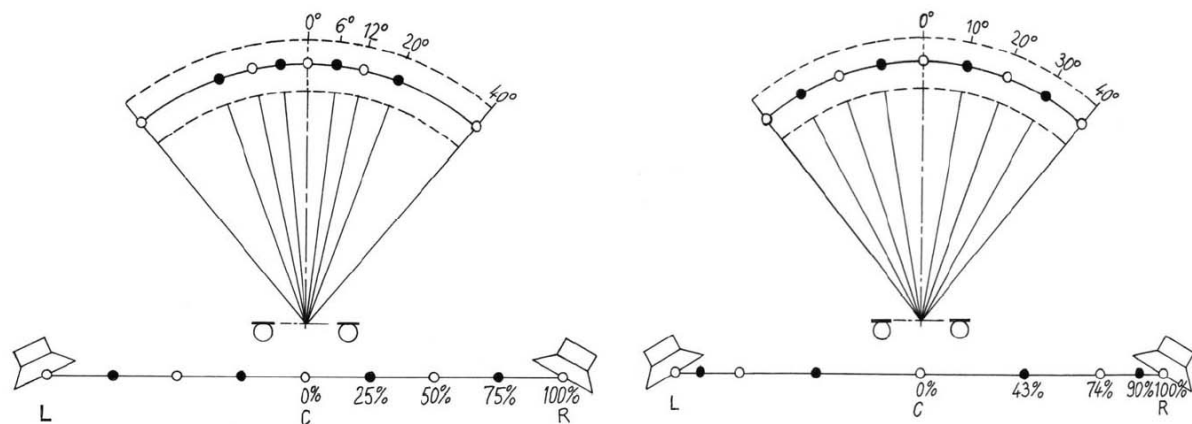


Abbildung 4-6 : Abbildungs-Nichtlinearitäten bei Laufzeitstereophonie: Lineare Abbildung bei der Wiedergabe bei ungleichförmig aufgestellten Schallquellen; nichtlineare Abbildung bei gleichförmiger Aufstellung (vgl. Sengpiel, Laufzeitstereophonie: Abbildungs- Nichtlinearitäten).

³² Genauere Schilderungen finden sich in Blauert, 1974, S. 178 ff.

³³ Eine Auslenkung von 0 % entspricht dabei der Mitte der Wegstrecke zwischen zwei Lautsprechern, sowie 100 % der Lokalisation in nur einem der beiden Lautsprecher. Die anderen Werte entsprechen einer linearen Einteilung der Wegstrecke zwischen 0 % und 100 %.(vgl. Sengpiel, Theoriegrundlagen: Laufzeitstereophonie)

Die Bewegung von Phantomschallquellen kann bei Laufzeitstereophonie durch eine Bewegung der realen Schallquelle zwischen den Mikrofonen realisiert. Eine nachträgliche Bearbeitung ist in der Regel nicht mehr möglich. Durch die wenig korrelierten Mikrophonsignale sind die Phantomschallquellen bei reiner Laufzeitstereophonie schwach ausgebildet, was zu einer größeren Lokalisationsunschärfe führt. Hinzu kommt, dass die meisten Hauptmikrophone nach dem Laufzeitstereophonieverfahren für Mehrkanalaufnahmen mit relativ großen Abständen der Mikrophone insbesondere zwischen vorderen und hinteren Kanälen arbeiten³⁴. Somit sind vor allem bei seitlichen und hinteren Schallquellen die Signale der einzelnen Mikrophone schwach korreliert.

Theile sagt dazu: „Ein reiner Laufzeitunterschied führt schon bei größeren interauralen Werten (im Bereich etwa 0,5 ... 0,8 ms) zu seltsamen Erscheinungen bezüglich Lokalisierbarkeit und Klangfarbe. Oberhalb etwa 1 ms wird der Effekt ziemlich katastrophal. [...] Insbesondere für größere Phantomschallquellenauslenkungen liefert eine geeignete Kombination gleichsinniger Laufzeit- plus Pegeldifferenzen bessere Ergebnisse bezüglich Lokalisationsschärfe und Klangfarbe, als reine Laufzeitdifferenzen“ (Theile: Surround Sound Forum Mails).

Eine Vorhersage über die Auswirkungen des Downmix bei Laufzeitstereophonie gestaltet sich sehr schwierig, da man im Gegensatz zur Intensitätsstereophonie keine genauen Angaben machen kann, auf welche Weise sich die Signale für den linken und rechten Audiokanal nach der Addition zusammensetzen werden.

Da keine Intensitätsunterschiede zwischen den Signalen der vorderen und hinteren Audiokanäle vorhanden sind, werden die angesprochenen technischen Fehler nicht auftreten. Das Audiosignal wird leiser wenn der Anteil der Surround-Kanäle am Gesamtklang größer wird.

Es ergeben sich jedoch durch die Addition des Centerkanals und der Surroundkanäle durch die Monosummenbildung Kammfiltereffekte. Diese entstehen bei Laufzeitstereophonie z.B. durch seitliche Einstrahlung der Schallquelle. In Abhängigkeit von der Mikrofonbasis a (siehe Abbildung 4-7) ergeben sich Phasendifferenzen, so dass der

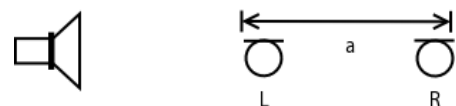


Abbildung 4-7 : Seitlicher Schalleinfall(vgl. Sengpiel, Welche Mikrofonbasis brauchen Raummikrophone 2)

³⁴ siehe z.B. Birkner Kapitel 3.2.2

Korrelationsgrad der Signale zwischen 1 und 0 hin- und herwechselt. Frequenzen deren Laufzeitverzögerung $\Delta t = a/c$ der halben Schwingungsperiode der Frequenz $f = 1/\Delta t$ entsprechen, werden ausgelöscht. Für die gleichphasigen Anteile ergeben sich dabei Maxima von +6 dB.

Dieses Phänomen ist bei der Wiedergabe über eine 2/0-Stereo-Standard-Aufstellung normalerweise nicht hörbar³⁵. Bei der elektrischen Addition zu einem Monosignal hingegen löschen sich die gegenphasigen Anteile aus und es kann zu Klangverfärbungen kommen (vgl. Sengpiel, Welche Mikrofonbasis brauchen Raummikrophone 2).

Je geringer also die Mikrofonbasis gewählt wird, desto geringer wird die Gefahr vom Kammfiltereffekten. Die Korrelation zwischen den beiden Mikrofonsignalen wird bei kleinen Mikrofonbasen größer. Nach Kapitel 4.3.2 trägt jedoch vor allem der seitliche, unkorrelierte Schall zur Räumlichkeit einer Aufnahme bei. Vor allem die tieffrequenten Anteile einer Aufnahme wirken durch eine kleine Mikrofonbasis während der Wiedergabe im Gegensatz zu den diffus klingenden Höhen eher nah und im Center positioniert, da die Korrelation für diese Signalanteile sehr hoch ist (vgl. ebd. 1). „Bass Enhance“ Algorithmen, wie beispielsweise in den Geräten der Firma Lexicon sorgen daher durch Phasenverschiebungen der tiefen Frequenzen um 90° für eine Dekorrelation der Signale und somit für eine Verstärkung des Räumlichkeitseindrucks (vgl. Griesinger: A Note to Sengpiel). So muss bei Laufzeitstereophonie immer ein Kompromiss zwischen der gewünschten Räumlichkeit und der Monokompatibilität gefunden werden.

Für die Abbildung von Phantomschallquelle zwischen den vorderen Lautsprechern wird das Klangbild enger (vgl. Wittek, Folie 17). Dies liegt daran, dass im Gegensatz zur 3/2-Stereo-Standard-Aufstellung sich die gleichen Signale auf einer geringeren „zeitlichen Basis“ sammeln, und so eine klangliche Dichte erzeugen. Dadurch werden Bewegungen in den vorderen Lautsprechern wesentlich schlechter lokalisierbar. Die Bewegungsrichtungen bleiben jedoch erhalten.

Seitliche Bewegungen von Phantomschallquellen dürften aufgrund der Tatsache, dass diese bei Laufzeitstereophonie im Gegensatz zur Intensitätsstereophonie noch instabiler sein werden, zwar möglich, aber schon in der 3/2-Stereo-Standard-Aufstellung schwer zu

³⁵ Hörbar werden Kammfiltereffekte jedoch durch den falschen Einsatz des Panoramareglers bei Laufzeitstereophonie. Erst ab einer Mikrofonbasis $a > 1\text{ m}$ sollten daher Aufnahmen nach dem Prinzip der Laufzeitstereophonie mittels Panpots „eingeeengt“ werden. (vgl. Sengpiel aus: G. Theile: Hauptmikrophone und Stützmikrophone – neue Gesichtspunkte für ein bewährtes Aufnahmeverfahren. Bericht zur 13. Tonmeistertagung 1984, S. 170-184)

lokalisieren sein. Die Folgen des Downmix sind dabei schwer einzuschätzen. Die durch die Monosumme entstehenden Kammfiltereffekte werden durch die 90° Phasenverschiebung der Surroundkanäle beim Dolby Pro Logic II und ITU-90° Downmix noch verstärkt. Sind genügend räumliche Anteile vorhanden, so kann die ehemals seitliche Bewegung nun als Hin- oder Wegbewegen wahrgenommen werden.

Der Dolby Pro Logic Downmix bringt zusätzliche Unvorhersehbarkeiten, da hier zusätzlich eine Monosumme der Surroundkanäle entsteht. Es ist anzunehmen, dass im Gegensatz zur Intensitätsstereophonie eine Bewegung von der Mitte zur Seite nur sehr schwer wahrzunehmen ist, da die zeitlichen Bezüge der Signale durch den Downmix zwar noch vorhanden sind, es jedoch wie beim Downmix der vorderen Audiokanäle zu einer Verdichtung der Audiosignale kommen wird.

Eine Bewegung von Phantomschallquellen zwischen den Surroundkanälen wird auch bei der Laufzeitstereophonie unter Umständen möglich sein. Als problematisch erweist sich dabei die meist große Mikrofonbasis der Surround-Mikrophone und den daraus resultierenden instabilen Phantomschallquellen bei vielen, der für Mehrkanalaufnahmen vorgeschlagenen Hauptmikrofon-Systemen. Bei geringeren Mikrofonbasen und damit auch geringeren Laufzeitunterschieden kann der Downmix gelingen, wobei auch hier ein Verdichtung des Klangbildes wahrscheinlich ist. Die Bewegungsrichtung kann unter günstigen Bedingungen erhalten bleiben.

Bei einem Dolby Pro Logic kompatiblen Downmix wird durch die Monosumme der Surroundkanäle die Bewegung vollständig verloren gehen, da die zeitlichen Bezüge der beiden Audiosignale nicht mehr in ihrer gewünschten Form erhalten bleiben.

Insgesamt lässt sich für die Realisation von bewegten Phantomschallquellen im 3/2-Stereobild festhalten, dass diese wesentlich schlechter zu lokalisieren sind und damit auch der Downmix-Kompatibilität wesentlich geringere Chancen eingeräumt werden. Der große Vorteil, der Laufzeitstereophonie eine gute Räumlichkeit zu gewährleisten, kann sich auf den Downmix evtl. positiv auswirken, wenn Bewegungen von Phantomschallquellen zwischen den vorderen und hinteren Lautsprechern im 2/0-Stereobild als Hin- oder Wegbewegen erhalten bleiben sollen.

4.3.3.3 Richtungsabbildung und Bewegung mit Hilfe von Äquivalenzstereophonie

Die Äquivalenzstereophonie wird ebenso wie die Laufzeitstereophonie in der Praxis eher in der Mikrophonaufnahmetechnik als in der Postproduktion eingesetzt.

Bei der Äquivalenzstereophonie sind Pegel- und Laufzeitdifferenzen gleichsinnig wirksam, so dass die Summe der Hörereignisrichtungen durch Pegel- und Laufzeitdifferenzen die Gesamthörereignisrichtung ergibt.

Für breitbandige Signale lassen sich bei guten Wiedergabebedingungen auf dem idealen Abhörplatz hinreichend genaue Aussagen über die Lokalisation der Hörereignisrichtung machen, wenn die Laufzeit- und Pegeldifferenz ($\Delta t, \Delta L$) von zwei Mikrophonesignalen bekannt sind. So ergibt sich die Gesamthörereignisrichtung zu:

$$b(\Delta L, \Delta t) = b_1(\Delta L) + b_2(\Delta t)^{36}.$$

(siehe Abbildung 4-8) (vgl. Sengpiel, Theoriegrundlagen Äquivalenz-Stereofonie).

Die Summe der Hörereignisrichtung ist dabei konstant, wenn sich Laufzeit- und Pegeldifferenzen gegenseitig ergänzen. Dabei ändert sich jeweils der Aufnahmebereich des Mikrophonsystems und die psycho-akustische Empfindung der Klangabbildung und der Räumlichkeit, je nachdem ob mehr Laufzeit- oder Pegeldifferenzen an der Hörereignisbildung beteiligt sind (vgl. ebd. Vergleich einiger Äquivalenzmikrophonsysteme).

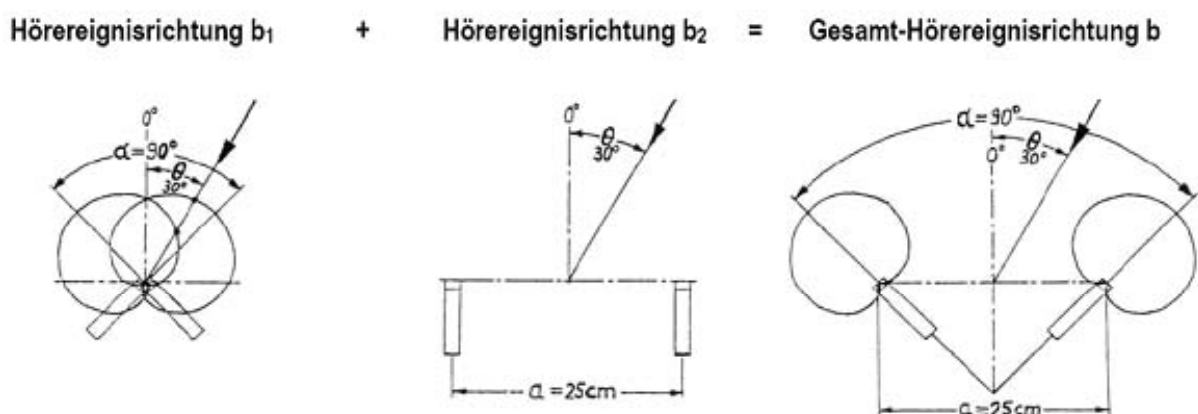


Abbildung 4-8 : EBS Mikrophonsystem (Sengpiel, Theoriegrundlagen: Äquivalenzstereophonie)³⁷

³⁶ b = Hörereignisrichtung auf der Stereo-Lautsprecherbasis in Prozent von der Stereomitte $C = 0\%$ aus Hörereignisrichtung 100% = Lokalisation aus der Richtung eines Lautsprechers. (vgl. ebd.)

³⁷ Hier im Bild am Beispiel des EBS-Mikrophonsystem mit Niere/ Niere, $\alpha = \pm 45^\circ = 90^\circ$ und $a = 25\text{cm}$ und Schalleinfallswinkel von $\theta = 30^\circ$. Mit θ = Schalleinfallswinkel von der 0° -Mitte des Mikrophonsystems aus für eine bestimmte Hörereignisrichtung; α = Achsenwinkel = Gesamtwinkel zwischen den Mikrofonhauptachsen; a = Mikrophonbasis = Abstand der Mikrophone voneinander (vgl. Sengpiel, Theoriegrundlagen Äquivalenz-Stereofonie)

Der Vorteil der Äquivalenzstereophonie liegt in der Kombination und dem gleichsinnigen und unterstützenden Zusammenwirken von Laufzeit- und Pegeldifferenzen und der damit verbundenen „natürlichen“ Klangwiedergabe: Das frühere Signal ist lauter, das spätere leiser.

Zu den Äquivalenzstereosystem gehören z.B. das NOS-System, das EBS-System, das RAI-System und das ORTF-System. In der Mehrkanalaudiotechnik gehört das sogenannte IRT-Kreuz (Abbildung 4-9) ebenfalls in diese Kategorie und kann prinzipiell als ein nach vorne und hinten gerichtetes Zweikanal-Mikrofon angesehen werden.

Äquivalenzstereophonie geht meistens einher mit einer guten Darstellung der Räumlichkeit sowie einer guten Lokalisation. Die Lokalisationsschärfe ist dabei besser als bei reiner Laufzeit- oder Intensitätsstereophonie. Gegensinnige Wirkungen der Mikrophonesignale führen nicht zu einer Subtraktion der Lokalisationswinkel sondern zu einer



Abbildung 4-9 : IRT-Kreuz mit vier Schoeps MK-4 Mikrofonen

Abnahme der Lokalisationsschärfe. Eine optimale Lokalisation wird demnach erreicht, wenn Laufzeit- und Pegeldifferenzen in etwa gleichen Anteilen zur Bildung des Hörereignis beitragen.

In der Praxis lassen sich die Anteile von Laufzeit- und Pegeldifferenzen zur Hörereignisbildung nicht eindeutig abschätzen, da zu viele Faktoren hier Einfluss nehmen: Mikrofontyp, Richtcharakteristik, Ausrichtung der Mikrophone, Mikrofonbasis, Abstand zur Schallquelle und Standort der Schallquelle (vgl. Dickreiter, Band1, S. 324).

Für den Downmix und die Darstellung von Bewegungen bedeutet der Einsatz von Äquivalenzstereophonie, dass die Vor- und Nachteile von Intensitäts- und Laufzeitstereophonie gleichermaßen zum Tragen kommen. Bei der verwendeten Anordnung überwiegen dabei diejenigen Parameter und Nachteile des für die Darstellung des Hörereignisses maßgeblichen Aufnahmeprinzips.

Die in Kapitel 4.3.1 angesprochenen technischen Fehler werden bei der Äquivalenzstereophonie je nach verwendetem System zwar auftreten, jedoch im Gegensatz zur reinen Intensitätsstereophonie geringere Auswirkungen haben.

Durch die geringen Mikrofonabstände und die damit verbundenen geringen Laufzeitunterschiede treten Kammfiltereffekte erst bei höheren Frequenzen auf. Somit sind die Auswirkungen der Monosummenbildung durch die Addition des Centerkanals und der Surroundkanäle zu den vorderen Audiokanälen nicht so gravierend wie bei reiner Laufzeitstereophonie. Zusätzlich können seitliche Bewegungen von Phantomschallquellen durch die Räumlichkeit der Aufnahmen im Gegensatz zu reinen Intensitätsstereophonie mit größerer Wahrscheinlichkeit als Hin- oder Wegbewegen erhalten bleiben.

Eine Nachbearbeitung von Aufnahmen, z.B. durch Pegelkorrekturen, ist nach meiner Erfahrung von Aufnahmen mit einem IRT-Kreuz sehr gut möglich, so dass sich gewünschte Effekte herausarbeiten lassen, ohne dabei große Klangverluste zu befürchten. Wichtig hierbei ist, dass die Nachbearbeitung nur die vorhandenen Klangeigenschaften unterstützt und nicht gegensinnig gearbeitet wird.

Insgesamt sind meiner Ansicht nach Aufnahmen nach dem Prinzip der Äquivalenzstereophonie für die Darstellung der Bewegung von Phantomschallquellen sowohl vor als auch nach dem Downmix besser geeignet als reine Laufzeit- oder Intensitätsstereophonie.

Zu beachten ist auch bei Äquivalenzstereophonie, dass durch die Phasenverschiebung von 90° bei der Dolby Pro Logic kompatiblen Dolby Digital Mischung die ursprünglichen Äquivalenzen gestört werden, so dass hier mit einer schlechteren Lokalisationschärfe bei der 3/2-Stereo-Wiedergabe gerechnet werden muss.

4.3.4 Aufteilung des Klangfeldes und Auswirkungen auf die Dramaturgie

Zu den erwähnten Folgen eines Downmix ergeben sich außerdem ästhetische Veränderungen durch die Reduktion der Wiedergabekanäle.

Die künstlerische Absicht durch die Positionierung im 3/2-Stereobild Hörereignissen eine bestimmte Aussage zu erteilen, kann durch den Downmix verloren gehen. Durch die neue Anordnung im 2/0-Stereobild verändern sich nicht nur die gewünschten Positionen der Hörereignisse, sondern auch ihre gegenseitigen räumlichen Bezüge. Zusammenhänge werden damit evtl. nicht mehr eindeutig dargestellt.

Das gewünschte Zusammenwirken der einzelnen Hörereignisse kann durch den Downmix verändert werden. Hat sich der Regisseur z.B. entschieden, ein Hörereignis außerhalb des Handlungsmittelpunktes zu anzuordnen, indem dieses in den Surround-Lautsprechern positioniert wurde, so wird dieses Hörereignis durch den Downmix evtl. direkt in die

Handlung integriert. Somit werden Bezüge zwischen Hörereignissen hergestellt, die ursprünglich nicht beabsichtigt waren. Die „Hörperspektive“ auf die einzelnen am Klang beteiligten Elemente verändert sich. Die Ausdehnung der einzelnen Schallquellen wird durch den Downmix verändert. Eine Atmosphäre, die im 3/2-Stereobild transparent und durchsichtig wirkt, so dass sich weitere Schallquellen gut in den Gesamtklang einfügen, kann nach dem Downmix zu wenig Raum für weitere Elemente lassen und durch die klangliche Dichte überladen wirken.

Das Timing der einzelnen Elemente in einem Klangbild kann durch den Downmix eine völlig andere Wirkung bekommen, so dass die ursprünglich gewünschten Effekte verloren gehen. Gerade bei der Präsentation von Audio-Programmen ohne Bild ist die zeitliche Abfolge der Hörereignisse für das Verständnis besonders wichtig. Im Gespräch mit unterschiedlichen Hörspielregisseuren äußerten diese, dass bei der Produktion von 3/2-Stereo oft ein anderer „Schnittrhythmus“ als bei 2/0-Stereo-Produktionen angewendet wird.

Die Sprachverständlichkeit kann durch die räumliche Nähe der einzelnen Hörereignisse, die durch den Downmix entstanden ist, geringer werden. Es muss daher schon bei der Produktion der 3/2-Stereo-Mischung darauf geachtet werden, an welcher Position im 2/0-Stereobild sich die einzelnen Elemente nach dem Downmix wiederfinden.

Dialoge, die sich zwischen den vorderen und hinteren Audiokanälen abspielen, können durch den Downmix zu unausgewogenen 2/0-Stereobildern führen, die zu rechts- oder linkslastig erscheinen.

Durch den Einsatz eines diskreten Centerkanals ist es im 3/2-Stereobild möglich durch die Platzierung von Dialogen eine hohe Sprachverständlichkeit zu erreichen, die nach dem Downmix nicht mehr in der gleichen Form vorhanden ist. So können die ursprünglichen Centersignale nach dem Downmix in Klangfarbe und Stabilität stark variieren. Die neue Phantomschallquelle wirkt dabei indirekter, da die Lokalisationsunschärfe für reale Quellen geringer ist, als für Phantomschallquellen (vgl. Theile, 2003, Folie 8, ff).

Bewegungen und Bewegungsrichtungen von Hörereignissen sind evtl. nicht mehr richtig nachvollziehbar. Es können wichtige Indikatoren zum Verständnis der Produktion fehlen und der Hörer kann somit das Wahrgenommene falsch interpretieren.

Die Lautstärkeverhältnisse der Hörereignisse können nicht mehr optimal aufeinander abgestimmt sein. Die Addition der Surroundkanäle zu den vorderen Audiokanälen kann bei sehr dominanten Signalen in den Surroundkanälen dazu führen, dass diese trotz der

Pegelabsenkung beim Downmix die vorderen Audiokanäle verdecken. Gerade Spezialeffekte oder auch Zuschauerreaktionen bei einem Live-Konzert können dazu führen, dass die wichtigen Informationen untergehen. Andererseits können wichtige Elemente in den Surroundkanälen durch eine Pegelreduktion nicht mehr gut genug wahrgenommen werden. Durch die Verdichtung des Klangbildes wird es in einigen Fällen zudem zu einer Veränderung der Gesamtdynamik kommen.

Diese Punkte sollten neben den technischen Überlegungen zum Downmix verstärkt eine Rolle spielen, da sie direkt in das künstlerische Schaffen eingreifen. Toningenieure und Regisseure sollten gleichermaßen die Folgen eines Downmix kennen und das Programm dementsprechend gestalten.

5 Hörvergleich

Ergänzend zu den theoretischen Überlegungen wird nun anhand eines praktischen Beispiels die Downmix-Kompatibilität von bewegten Signalen untersucht. Mit Hilfe eines Hörvergleichs soll festgestellt werden, ob sich eine Variante des Downmix besonders hervorhebt. Zu diesem Zweck wurde mit dem Hörspielregisseur Simon Jaspersen ein Kurzhörspiel produziert und dieses in die verschiedenen Formate kodiert.

5.1 Produktionsbericht: Ein Kurzhörspiel in 3/2-Stereo – „3 x Unschuld“

Mit der Produktion eines Kurzhörspiels konnte am praktischen Beispiel mit verschiedenen Bewegungen im Klangbild experimentiert werden, um die Downmix-Kompatibilität einer solchen Produktion zu untersuchen. Die Vorbereitungsphase der Produktion fand vom 15. – 19. August 2005, die Aufnahmen und Post-Produktion in der Zeit vom 22. – 26. August 2005 statt.

5.1.1 Konzeption und Planung

Das Genre Hörspiel eignet sich besonders gut für Experimente mit Klangbildern, da hier im Gegensatz zum Film, eine Reduzierung auf den Hörsinn stattfindet und sich der Hörer somit besser auf ungewohnte Hörerlebnisse konzentrieren kann. Das Genre lässt viele gestalterische Freiräume und legt die Hörperspektive nicht schon durch einen Bildausschnitt fest.

Der freie Hörspielregisseur Simon Jaspersen konnte für die Produktion gewonnen werden und erklärte sich bereit, das Hörspielskript zu verfassen³⁸ und die Auswahl der Sprecher zu übernehmen.

Als Spielort wurde ein U-Bahnhof ausgewählt, da zu erwarten war, dass sich durch die an- und abfahrenden Züge die gewünschten Bewegungen im Hörbild einfach realisieren lassen. Die gewünschten Bewegungen: Von hinten links nach hinten rechts“ (LS→RS) (oder umgekehrt), einseitig von vorne nach hinten (L→LS, R→RS) und einseitig von hinten nach vorne (LS→L, RS→R) wurden dem Regisseur mitgeteilt und von ihm in das Skript eingearbeitet.

Die Hörereignisse wurde so gewählt, dass zu erwarten war, dass durch den Downmix besondere Probleme entstehen würden. Eine Bewegung vorne links nach rechts (oder umgekehrt) schien nicht von besonderer Bedeutung, da hier nach dem Downmix keine wesentlichen Veränderungen zu erwarten sind.

Zusätzlich sollten einige Elemente des Hörspiels so im Hörbild platziert werden, dass sie sich von der klassischen Positionierung im 3/2-Stereobild abheben. Das Experimentieren mit unterschiedlicher Positionierung von Hörereignissen sollte zeigen, inwieweit sich das Surround-Format eignet, neue Höreindrücke zu schaffen, und welche Auswirkungen diese Positionen auf den Downmix haben.

5.1.2 Vorüberlegungen für die Durchführung der Produktion

Um eine möglichst hohe Sprachverständlichkeit zu gewährleisten und die Produktion einfach zu gestalten, entschied ich mich Surround-Atmo-Aufnahmen auf Hamburger U-Bahnsteigen zu machen, um diese dann mit den Aufnahmen der Sprecher im Studio zu mischen.

Die Anforderungen an die Technik für die Atmo-Aufnahmen wurden wie folgt festgelegt:

- Digitale Aufzeichnung von mindestens vier Spuren
- Phantom Power für den Einsatz von hochwertigen Kondensatormikrofonen
- tragbar und flexibel einsetzbar / leichter Transport
- Möglichkeiten zur Befestigung von Windschützen an den Mikrofonen
- möglichst unabhängig von Stromquellen

³⁸ Das Hörspielskript befindet sich im Anhang.

Von den bekannten Mehrkanal-Mikrophonierungsverfahren schienen aufgrund der Anforderungen an Größe und Flexibilität das Doppel-MS-Verfahren und das IRT-Kreuz am besten geeignet, um die Aufnahmen in der U-Bahn zu realisieren. Zusätzlich zur guten Lokalisation der U-Bahn war auch ein realistisches Hörbild des U-Bahnsteigs gewollt und so entschied ich mich für den Einsatz des IRT-Kreuz. Da dieses Mikrophonsystems auf dem Prinzip der Äquivalenzstereophonie beruht, würde eine gute räumliche Auflösung und eine gute Lokalisation gewährleistet sein. Das IRT-Kreuz wurde mit zwei Stereo-Schienen und vier Schoeps MK4 Mikrofonen realisiert, die auf einem einfachen Mikrofonständer montiert wurden. Die Kantenlänge zwischen den einzelnen Mikrofonen betrug wie für Nierenmikrophone empfohlen 25 cm (vgl. Theile, 2001).

Vier einfache Schaumstoffwindschütze wurden auf die Mikrofone aufgesteckt, da leider keine Windkörbe mit Fell vorhanden waren. Diese Konstruktion war leicht zu transportieren und ließ sich daher auch am Ort der Aufnahmen gut verstellen.

Das IRT-Kreuz wurde ursprünglich von Günter Theile als Ergänzung zu einem L-C-R-Hauptmikrofon entworfen (vgl. ebd.). Diese Konstruktion eignet sich nach meiner Erfahrung auch hervorragend zur Aufnahme von Atmos. Wenn diese Aufnahmen dann in der Postproduktion mit anderen 3/2-Stereoaufnahmen gemischt werden, ergibt sich ein insgesamt natürliches Klangbild.

Die Aufnahmen wurden mit einem Digidesign „Digi002“ Audiointerface und Laptop mit der Audiosoftware NUENDO der Firma Steinberg durchgeführt. Das Digidesign „Digi002“ lieferte die nötige Spannungsversorgung für die Mikrofone und konnte auch eine Aufnahmequalität von 24 Bit bei 44,1 kHz Abtastrate bereitstellen. Zusätzlich stellt das Digidesign „Digi002“ für jeden Mikrofon-Eingang einen Hochpass (Trittschallfilter) zur Verfügung, so dass hier ein zusätzlicher Schutz gegen Übersteuern durch Windgeräusche gegeben war. Eine Abhörmöglichkeit vor Ort existierte in Form eines Kopfhörers, mit dem jedoch aufgrund der hohen Lautstärke der Bahn keine gültigen Aussagen über die Signalqualität gemacht werden konnte. Es musste sich somit auf die optische „Clipping“ Anzeige in NUENDO verlassen werden.

Da die Elektrizität auf dem U-Bahnsteig nach Absprache mit der Hamburger Hochbahn AG genutzt werden konnte, stellte diese Kombination von Equipment den besten Kompromiss bezüglich der formulierten Anforderungen dar.

5.1.3 Produktionsablauf

Am Montag den 22. August 2005 um 5 Uhr morgens konnten die Aufnahmen in der U-Bahn auf den Bahnsteig der Haltestelle Jungfernstieg gemacht werden. Um diese Uhrzeit war mit den wenigsten Störungen durch Fahrgäste zu rechnen.

Die räumlichen Aufteilungen der Haltestelle und die An- und Abfahrrichtungen entsprachen unseren Vorstellungen. In der Mitte des Bahnsteigs befand sich eine Rolltreppe, die als „akustischer Farbtupfer“ dienen sollte.

Für die Aufnahmen wurden zwei verschiedene Standorte ausprobiert (siehe Anlage: Lageplan U-Bahn), wobei Standpunkt B geeigneter schien, da hier die Bahn größere Eigengeräusche produzierte.

Zur besseren Organisation der Aufnahmen wurden die Mikrophone markiert und wie folgt benannt: Spur 1= links; Spur 2 = rechts; Spur 3 = links hinten; Spur 4 = rechts hinten.

Die Bahnen wurden in vier Variationen aufgenommen:

- I. links / rechts auf Bahn 1 gerichtet, Ohrhöhe; mit Hochpass; Standpunkt A
- II. links / recht auf Bahn 1 gerichtet, Ohrhöhe; mit Hochpass; Standpunkt B
- III. links / recht auf Bahn 1 gerichtet, ca. 2, 30 m Höhe; mit Hochpass; Standpunkt B
- IV. links / recht auf Bahn 1 gerichtet, ca. 2, 30 m Höhe; ohne Hochpass; Standpunkt B

Beim späteren Abhören musste leider festgestellt werden, dass die Aufnahmen zum großen Teil durch Windstöße gestört waren. Windkörbe mit Fell wären für die Aufnahmesituation besser geeignet gewesen. Der Hochpass erwies sich als essentiell, da die Aufnahme IV. (ohne Hochpass) sehr starke Störungen durch Windgeräusche aufwies.

Die Aufnahmen I. und II. waren im wesentlichen brauchbar, wobei Aufnahme I. im Gegensatz zur Aufnahme II. die wenigsten Problem mit Windgeräuschen aufwies. Dafür war die Abbildung der Bahngeräusche bei Aufnahme II. wesentlich „plastischer“, so dass für die Produktion diese Aufnahme ausgewählt wurde. Die Aufnahmen III. und IV. waren zum Teil stark bis sehr stark durch Windgeräusche gestört, so dass diese in der Produktion keine Verwendung finden konnten. Die beste Abbildung der Geräuschkulisse lieferten die Aufnahmen, die nah an der Bahn in Ohr-Höhe gemacht wurden.

Zusätzlich wurden in der U-Bahnhaltestelle noch einzelne Aufnahmen von Flaschenklappern, der Rolltreppe und den Zügen gemacht, die aber bis auf das Flaschenklappern keine Verwendung in der Produktion fanden, da die Aufnahmen mit dem IRT-Kreuz voll ausreichend waren.

Später wurden die einzelnen Spuren mit NUENDO so im Hörbild angeordnet, dass der jeweils gewünschte Szenenaufbau erreicht wurde. Um die Bewegung der Bahnen zu unterstützen, wurden die einzelnen Spuren entsprechend der Einfahrriechung in den Lautstärken angeglichen³⁹. Die Windgeräusche konnten aus den Aufnahmen entfernt werden, ohne den Klangeindruck zu verfälschen.

Um die Sprachverständlichkeit zu gewährleisten, musste zusätzlich die sehr große Dynamik der Aufnahmen eingeschränkt werden, so dass der Pegel der einfahrenden Bahnen zum Teil sehr stark abgesenkt wurde. Insgesamt lässt sich feststellen, dass die Bearbeitung von Mehrkanal-Audiomaterial (Schnitt, Dynamik) sehr viel einfacher zu handhaben ist, als bei Zweikanalproduktionen. „Fehler“ werden wesentlich besser verdeckt.

Die Schauspieler wurden mit einem Neumann U-87 unabhängig von einander aufgenommen und entsprechend im Hörbild platziert. Für die Monologe der einzelnen



Abbildung 5-2 : „Wachmann“
Oliver Törner

Szenen hatte sich der Regisseur eine Position im Hörbild gewünscht, die sich deutlich von den Szenen abhebt und die Subjektive der Sprecher betont. Die Monologe wurden daher auf einer Position in der Mitte des Hörbildes platziert, indem sie aus allen Lautsprechern mit gleicher Lautstärke wieder-gegeben wurden. Durch die Positionierung der Monologe sollte eine große Lokalisationsunschärfe (vgl. Blauert, S. 30) erreicht werden, um dem Hörer das Gefühl zu geben, das Hörereignis würde aus dem „Nichts“ kommen, so dass der Unterschied zu den Dialogen in der Szene deutlich wird.



Abbildung 5-1 : Atmo- Aufnahme

³⁹ Wenn z.B. die Bahn eine Bewegung von vorne links nach hinten links machen sollte, wurde zuerst der Audiokanal vorne links lauter eingestellt, und dann der Audiokanal hinten links.

In den Szenen wurden die Sprecher in den vorderen Audiokanälen positioniert, um die Spielsituation herauszuheben. Mit Hilfe des „True Verb“ Hallprogramm der Firma Waves und des Surround Halls der Firma Steinberg wurden sie räumlich an die U-Bahn-Atmo angepasst. Eine Schwierigkeit hierbei war es, den Anteil der diffusen Raumanteile in den Surroundkanälen so anzupassen, dass ein räumlicher Höreindruck entsteht, ohne dass bei einer Sitzposition des Hörers hinter dem Sweet-Spot die Spielszene in den hinteren Lautsprechern wahrgenommen wird und das Hörbild somit „kippt“. Um die Sprachverständlichkeit zu erhöhen, wurde die Dynamik aller Sprecher mit dem Kompressor C1 von Waves bearbeitet.

Zusätzlich zu den bereits in der U-Bahn aufgenommenen Geräuschen (Flaschenklappern, Rolltreppe) wurden das Zerschlagen einer Glasflasche (im Leinenbeutel auf einem Stein) und unterschiedliche Schritte aufgenommen, die später in die Produktion integriert wurden.

Die Produktion wurde im Tonstudio der HAW gemischt und anschließend auf unterschiedlichen Abhörmöglichkeiten gegengehört. Zu diesem Zweck wurden eine DVD-Audio sowie eine DVD-Video erstellt. Diese konnten dann auf unterschiedlichen Konsumer-Anlagen sowie im „Mastering & Surround Factory“ Studio der Firma Boogie Park in Hamburg angehört werden.

5.2 Durchführung des Hörvergleichs

Das Ziel des Hörvergleichs war, festzustellen, welche Auswirkungen der Downmix auf die 3/2-Stereo-Produktion hat und ob die Qualitätsunterschiede von den Hörern wahrgenommen werden. Untersucht wurde Audiomaterial mit Bewegungen von Phantomschallquellen zwischen den Surround-Lautsprechern sowie zwischen den vorderen Audiokanälen und den Surround-Lautsprechern.

Ergänzend sollte herausgefunden werden, ob eine Unterscheidung zwischen der diskreten Wiedergabe von 3/2-Stereosignalen und 3/2-Stereosignalen mit einer Phasenverschiebung der Surroundkanäle von 90°, wie sie bei Dolby Pro Logic kompatiblen Dolby Digital Signalen auftritt, möglich ist.

5.2.1 Audiobeispiele für den Hörvergleich

Als Audiomaterial für den Vergleich wurden Ausschnitte aus dem Hörspiel verwendet, das als fünf diskrete 24 Bit/ 44,1 kHz-PCM-Audiofiles vorlag und mit verschiedenen Verfahren auf 2/0-Stereo automatisch heruntergemischt wurde. Zusätzlich wurde eine händische⁴⁰ 2/0-Stereo-Mischung angefertigt, die so angelegt wurde, dass die Positionen der Sprecher den Positionen nach dem Downmix entsprechen und dass das Timing der einzelnen Klangelemente übereinstimmt. Bis auf dieses Zugeständnis an den Downmix wurde die händische Version als eigenständige 2/0-Stereofassung gemischt. Alle Audiobeispiele wurden im Gesamtpegel so angepasst, dass der Pegel der Dialoge übereinstimmt.

Der Hörvergleich bestand aus drei unterschiedlichen Audiobeispielen⁴¹ die den Probanden in unterschiedlichen Versionen dargeboten wurden:

- Beispiel 1: Ausschnitt aus der Szene 1 des Hörspiels, der eine bewegte Phantomschallquelle von hinten links nach hinten rechts beinhaltet - 0,29 min.
- Beispiel 2: Ausschnitt aus der Szene 2 des Hörspiels, der eine bewegte Phantomschallquelle von vorne links nach hinten links beinhaltet – 0,35 min.
- Beispiel 3: Ausschnitt aus der Szene 3 des Hörspiels, der eine bewegte Phantomschallquelle von hinten rechts nach vorne rechts sowie von vorne links nach hinten links beinhaltet – 1,10 min.

Da die Audiobeispiele Ausschnitte aus verschiedenen Szenen des Hörspiels „3 x Unschuld“ waren, beinhalteten diese neben den Bahngeräuschen (= bewegte Phantomschallquelle) auch unterschiedliche Dialoge und Geräusche.

- Beispiel 1: Bahnfahrt von hinten links nach hinten rechts; zwei männliche Sprecher aus dem Off sowie in der Szene; unterschiedliche Geräusche, wie z.B. das Zerschlagen einer Glasflasche, Kampfgeräusche und Schritte.
- Beispiel 2: Bahnfahrt von vorne links nach hinten links; Stimme von einer Frau aus dem Off sowie in der Szene; eine Kinderstimme in der Szene; Schritte; Geräusch „Anrempeln“.

⁴⁰ Hiermit übernehme ich den von den österreichischen Toningenieuren geprägten Begriff für eine „handgemachte“ Stereomischung, da der Begriff meiner Ansicht nach den Sachverhalt anschaulich widerspiegelt.

⁴¹ Der Terminus „Audiobeispiel“ bezeichnet einen Ausschnitt aus einer Szene des Hörspiels, unabhängig vom Audioformat.

- Beispiel 3: Bahnfahrt von hinten rechts nach vorne rechts sowie von vorne links nach hinten links; männlicher Sprecher aus dem Off sowie in der Szene, Lautsprecherdurchsage hinten rechts; Schritte von hinten links nach vorne in die Mitte; Abschuss einer „Taser“- Waffe.

Die Audiobeispiele lagen in den folgenden Format-Versionen vor:

- Downmix nach ITU-R 775 mit Downmix-Koeffizient $k=0,7$
- Downmix nach ITU-R 775 mit Downmix-Koeffizient $k=0,7$ aus Dolby Pro Logic kompatiblen diskreten Audiospuren – ITU-90⁴²
- Downmix nach Dolby Pro Logic
- Downmix nach Dolby Pro Logic II
- Händische Version
- Diskrete 3/2-Stereo-Version
- Diskrete 3/2-Stereo-Version mit 90° Phasenverschiebung der Surroundkanäle

Bis auf die Dolby Pro Logic II Version konnten alle Downmixe mit der Audiosoftware NUENDO erstellt werden. Der Downmix nach ITU-R 775/ ITU-90° wurde mit Hilfe des NUENDO-Plugins „Mixconvert“ erstellt. Der Dolby Pro Logic Mix wurde mit dem NUENDO-Plugin „Mixconvert“ und einem digitalen Bandpassfilter der Firma Waves nach dem in Kapitel 3.1.1.1 vorgestellten Schema erstellt.

Der Dolby Pro Logic II Downmix wurde freundlicherweise von der Firma Dolby in England auf einem Dolby DP563 Dolby Surround und Pro Logic II Referenz Encoder erstellt, da in Deutschland keines dieser Geräte zur Verfügung stand. Die Codierung konnte aus zeitlichen Gründen nur einmal ausgeführt werden; evtl. nötige Anpassungen waren daher nicht möglich und so wurde die Kodierung mit den Standardparametern durchgeführt. Das Audiobeispiel weist leider teilweise wahrnehmbare Verzerrungen auf, da das Dolby Pro Logic II Signal zu hoch angesteuert war. Es wurde versucht, für den Hörvergleich Material ohne Verzerrungen zu verwenden.

Die diskrete 3/2-Stereo-Version war die Originalmischung des Hörspiels. Die diskrete 3/2-Stereo-Version mit 90° Phasenverschiebung der Surroundkanäle wurde aus der Originalmischung mit Hilfe des NUENDO Plugins „Mixconvert“ erstellt.

Die Audiobeispiele befinden sich auf der dieser Diplomarbeit beiliegenden DVD.

⁴² Siehe Kapitel 4.1

5.2.2 Versuchsbedingungen

Der Hörvergleich fand am 02. Dezember 2005 und am 05. Dezember 2005 im Raum 124 des Fachbereichs statt. Es haben insgesamt 17 Personen am Hörversuch teilgenommen. Am 02. Dezember wurden 9 Personen in eine Vierer-, eine Dreier- und eine Zweiergruppe aufgeteilt. Der Test am 05. Dezember wurde mit vier Zweiergruppen durchgeführt.

Der Raum wurde zur Verbesserung der Akustik an der Rückwand mit einem großen Stück Molton abgehängt, wodurch das Entstehen eines Flatterechos vermieden wurde. Die mit dem MLSSA™ Audiomeßsystem gemessene Nachhallzeit betrug 0,475 Sekunden und entsprach damit den Anforderungen an eine gute Abhörsituation. Der Aufbau bestand aus fünf aktiven „Genelec 1030“ Lautsprechern, die nach ITU-R Empfehlung Bs.775-1 auf Lautsprecherstativen in einer Höhe von 1,2 m aufgestellt waren. Die Lautsprecher wurden so angeordnet, dass diese mindestens 1m entfernt von den Begrenzungsflächen des Raumes aufgestellt werden konnten. Daraus ergab sich für die Lautsprecher-Aufstellung ein Radius von 2,5 m. Die Surround-Lautsprecher wurden so aufgestellt, dass sie mit einer Geraden zum Abhörpunkt einen Winkel von 110° zur Geraden des Center-Lautsprechers bildeten.

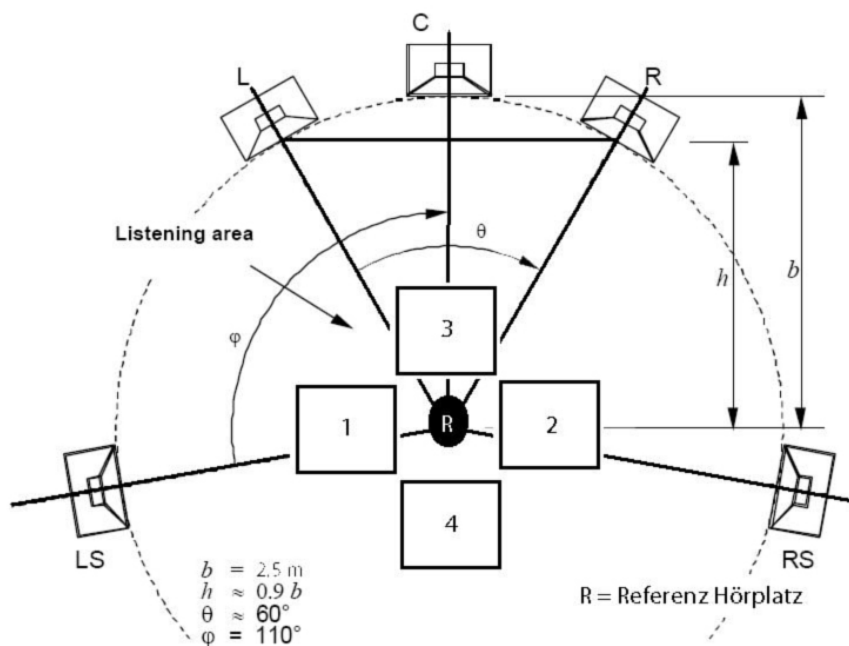


Abbildung 5-3 : Sitzpositionen der Probanden beim Hörvergleich

Der Wiedergabepegel wurde für die 2/0-Stereo- sowie für die 3/2-Stereo-Abhörsituation jeweils mit einem rosa Rauschen von -9 dBFS⁴³ auf einen Summenpegel von 82 dB SPL(A) am Referenzsitzplatz angepasst. Die Messung wurde mit dem Messgerät der Firma Brüel & Kjaer „2237 Controler“ in der Einstellung einer A-Bewertung und dem in der Software NUENDO vorhandenen Testsignal-Generator durchgeführt.

Als Zuspierer für den Hörversuch wurde ein Notebook mit der Audiosoftware NUENDO und dem Digidesign „Digi002“ Audiointerface verwendet. Der Pegel der einzelnen Lautsprecher wurde direkt in NUENDO angeglichen. Für jeden Abschnitt des Hörvergleichs wurde ein eigenes Projekt angelegt, so dass gewährleistet war, dass immer die optimale Abhörlautstärke eingestellt war.

Die Sitzposition der Probanden orientiert sich an der EBU Tech 3276-E (siehe Abbildung 5-3). Bei einer festgelegten Gruppengröße von maximal 4 Personen pro Gruppe wurden die Sitzplätze um den „Sweet Spot“ herum angeordnet. Bei kleineren Gruppen (mindestens zwei Hörer) wurden zunächst die Plätze 1 und 2 und dann Position 3 belegt. Der Sitzplatz der Probanden wurde auf dem Auswertungsbogen notiert.

Die Gruppe der Probanden setzte sich aus Personengruppen zusammen, bei denen zu vermuten ist, dass diese über ein gutes Gehör verfügen. Unter den Probanden waren neben Medientechnikstudenten, die sich auf die Audiotechnik spezialisieren, auch freie Toningenieure und Musiker. Eine Beschränkung auf diese Personengruppe war meiner Ansicht nach notwendig, da die geringen Unterschiede für ungeübte Hörer nur sehr schwer wahrzunehmen sind.



Abbildung 5-4 : Aufbau für den Hörvergleich

⁴³ Relativ zum digitalen Clip-Pegel 0dBFS

5.2.3 Versuchsbeschreibung und -durchführung

Für den Hörvergleich wurden vier verschiedenen Testreihen ausgearbeitet. Die erste, zweite und vierte Testreihe wurde dabei als ABX-Test durchgeführt. Diese Testreihen sollten feststellen, ob die Probanden zwischen den einzelnen Versionen der Audiobeispiele Unterschiede wahrnehmen können. Die dritte Testreihe sollte einzelne Klangmerkmale der unterschiedlichen 2/0-Stereo-Versionen der Audiobeispiele abfragen. Da die Audiobeispiele aus dem Kurzhörspiel „3 x Unschuld“ stammten, wurde den Probanden vor Versuchsbeginn die Handlung des Hörspiels erläutert, damit diese sich ganz auf die klanglichen Aspekte der Hörbeispiele⁴⁴ konzentrieren konnten.

In den Testreihen eins und zwei wurde den Testpersonen zuerst ein Referenz-Hörbeispiel vorgestellt. Anschließend wurden zwei weitere Hörbeispiele (Hörbeispiel 1, Hörbeispiel 2) vorgespielt, von denen eines dem Referenzbeispiel entsprach. Die Testhörer sollten im Anschluss auf dem Testbogen das Beispiel (Hörbeispiel 1, Hörbeispiel 2) ankreuzen, welches ihrer Ansicht nach dem Referenzbeispiel entspricht.

Testreihe 1 wurde mit dem Audiobeispiel 1, Testreihe 2 mit dem Audiobeispiel 2 durchgeführt. In beiden Testreihen wurden die folgenden Kombinationen von Formaten miteinander verglichen:

- a. Dolby Pro Logic II vs. händische Version
- b. ITU vs. händische Version
- c. ITU vs. Dolby Pro Logic
- d. Itu-90° vs. ITU

Die Reihenfolge der Vergleichspaare (a, b, c, d), sowie die Auswahl welches der beiden Hörbeispiele als Referenz, welches als Hörbeispiel 1 bzw. Hörbeispiel 2 vorgestellt wurde, wurde mit jeder Testgruppe verändert, um Fehler durch „Einhöreffekte“ zu vermeiden. Jede Gruppe von Testhörern bekam für jeden Testabschnitt die Hörbeispiele einmal vorgestellt und sollte im Anschluss eine Entscheidung treffen.

Zusätzlich hatten die Probanden die Möglichkeit in einer Prinzipskizze einer Surround-Lautsprecheranordnung durch Pfeile die ihrer Ansicht nach ursprüngliche Bewegungsrichtung der Phantomschallquellen einzuzeichnen.

⁴⁴ Hörbeispiele sind die unterschiedlichen Versionen der Audiobeispiele; z.B. Audiobeispiel 1 in der Dolby® Pro Logic II Version.

In Testreihe 3 sollten die Hörer auf der Grundlage von unterschiedlichen Parametern einzelne Hörbeispiele bewerten. Für diesen Test wurden Variationen des Audiobeispiels 3 verwendet:

- Downmix nach ITU-R 775 mit Downmix-Koeffizient $k=0,7$
- Downmix nach ITU-R 775 mit Downmix-Koeffizient $k=0,7$ aus Dolby Pro Logic kompatiblen diskreten Audiospuren – ITU-90⁴⁵
- Downmix nach Dolby Pro Logic
- Downmix nach Dolby Pro Logic II
- Händische Version

Wiederum wurde die Reihenfolge der Hörbeispiele mit jeder Testgruppe verändert, um Fehler durch „Einhöreffekte“ zu vermeiden. Jede Gruppe von Testhörern bekam für jeden Testabschnitt das jeweilige Hörbeispiel einmal vorgestellt und sollte im Anschluss eine Entscheidung treffen.

Die Parameter für die Bewertung stammten aus der EBU-Tech. 3286–E “Assessment methods for the subjective evaluation of the quality of sound programme material – Music” und wurden den Probanden vor dem Hörvergleich in deutscher Übersetzung ausgehändigt, damit diese sich mit den Parametern vertraut machen konnten. Die einzelnen Kategorien sollten anhand der Bewertungsskala für subjektive Einschätzungen nach EBU-Tech 3286 (1997) eingeteilt werden⁴⁶.

Mit der Testreihe 3 sollte herausgefunden werden, ob sich für die einzelnen Audiobeispiele eine Tendenz bezüglich der einzelnen Parameter erkennen lässt, anhand der die Ergebnisse der Testreihen 1 und 2 besser interpretiert werden können.

Im Audiobeispiel 3 ist im Gegensatz zu den anderen Audiobeispielen jeweils eine Bahn links und rechts zu hören; hierdurch sollte vermieden werden, dass der Downmix als zu links- oder rechtslastig empfunden wird.

In Testreihe 4 wurde den Probanden wie in den Testreihen 1 und 2 zuerst ein Referenz-Hörbeispiel vorgestellt. Anschließend wurden zwei weitere Hörbeispiele (Hörbeispiel 1, Hörbeispiel 2) vorgespielt, von denen eines dem Referenzbeispiel entsprach. Die Testhörer sollten im Anschluss auf dem Testbogen das Beispiel (Hörbeispiel 1, Hörbeispiel 2) ankreuzen, welches ihrer Ansicht nach dem Referenzbeispiel entspricht.

⁴⁵ siehe Kapitel 4.1

⁴⁶ siehe Anhang „Bewertungsskala für subjektive Einschätzungen“

Untersucht wurde hierbei Audiobeispiel 4 in der diskreten 3/2-Stereo-Version und der diskreten 3/2-Stereo-Version mit 90° Phasenverschiebung der Surroundkanäle.

Die Auswahl, welches der beiden Hörbeispiele als Referenz, welches als Hörbeispiel 1 oder Hörbeispiel 2 vorgestellt wurde, wurde mit jeder Testgruppe verändert. Jede Gruppe von Testhörern bekam für jeden Testabschnitt die Hörbeispiele einmal vorgestellt und sollte im Anschluss eine Entscheidung treffen.

Auf dem Testbogen war zudem noch ein Testteil b angefügt, der jedoch nicht existierte. Diese Fragestellung diente nur dazu die Konzentration und Aufmerksamkeit zu erhalten. Es hat sich in Versuchen der FH Düsseldorf herausgestellt, dass bei Nichtanfügen einer zusätzlichen Frage die letzten Fragen zu Hörversuchen nicht mehr konzentriert beantwortet wurden (vgl. Adam, S.66).

Die Original-Fragebögen befinden sich im Anhang der Arbeit.

5.3 Auswertung der Testreihen

5.3.1 Testreihe 1 und 2

Die einzelnen Bewertungsbögen wurden manuell erfasst und, nach Sitzposition geordnet, für jeden Vergleich und Testreihe ausgezählt, um festzustellen, ob hier signifikante Unterschiede in der Stimmverteilung zu bemerken waren. Dieses war jedoch nicht der Fall. Hatten die Probanden bei einem bestimmten Parameter keine Bewertung abgegeben, wurde der Parameter unter der Kategorie falsch erfasst.

Betrachtet man für die *Testreihe 1* den Vergleich (a) zwischen der händischen und der Dolby Pro Logic II Version, kann man feststellen, dass 82 % der Probanden das Referenzbeispiel richtig zuordnen konnten. Daraus folgt, dass bei dieser Kombination von Hörbeispielen ein deutlicher Unterschied wahrzunehmen war.

Im Vergleich klingt die händische Version meiner Ansicht nach wesentlich transparenter und ausgeglichener. Die Stimmen und Geräusche sind gut zu lokalisieren und lassen sich gut zuordnen. Der Raum wirkt homogen und die Tiefenstaffelung ist gut wahrnehmbar. Die Dolby Pro Logic II Version hingegen wirkt eher flach und dumpf. Eine deutliche Klangfärbung ist festzustellen sowie die Tendenz, dass die einzelnen Klangelemente untereinander „verwischen“ und sich räumlich nicht voneinander trennen lassen. Besonders deutlich wird dieses bei der Monologstimme. Der Downmix der Bewegung der Bahn ist zwar der Theorie entsprechend erfolgt, wirkt aber im Gegensatz zur händischen Version „verwaschen“ und undifferenziert. Auch die links-rechts-Bewegung ist nicht

eindeutig. Die Bahn scheint zwar von links einzufahren, dann lässt sie sich allerdings nicht mehr eindeutig in der Richtung bestimmen. Die Ursache dafür liegt in der in Kapitel 4.3 angesprochenen Problematik, dass beim Dolby Pro Logic II Downmix keine eindeutige Kanaltrennung der Surroundkanäle nach dem Downmix mehr vorhanden ist. Da sich im rechten Audiokanal des Lt/Rt-Signals nun auch Bahngeräusche befinden, die ursprünglich nur der linken Seite zugeordnet waren, addieren sich diese mit den Bahngeräuschen, die sich im rechten Audiokanal der 3/2-Stereo-Version befinden. Durch diese Addition sind nun die Bahngeräusche im rechten Audiokanal teilweise zu stark wahrnehmbar.

Beim Vergleich (b) zwischen der händischen und der ITU Version ist der Unterschied nicht mehr so deutlich festzustellen. Immerhin 71 % der Probanden konnten das Referenzbeispiel richtig zuordnen. Im Gegensatz zur Dolby Pro Logic II Mischung sind die Klangfärbungen nicht so deutlich wahrnehmbar und der Raumeindruck entspricht mehr der händischen Version. Allerdings lässt sich auch hier feststellen, dass sich die Stimme für den Monolog nicht angemessen in den Gesamtklang einfügt. Der Downmix der Bahn allerdings gelingt gut. Die Bewegung der Bahn lässt sich deutlich nachvollziehen und unterscheidet sich im Gegensatz zur händischen Version nur durch die geringere Räumlichkeit.

Das Ergebnis des Vergleich (c) war überraschend. Die Probanden konnten den Unterschied zwischen dem ITU-Downmix und der Dolby Pro Logic Version nicht eindeutig erkennen. Nur 59 % der Probanden war es gelungen, das Referenzbeispiel zu erkennen. Durch die Monosumme der Surroundkanäle beim Dolby Pro Logic Downmix waren diese beiden Hörbeispiele leicht auseinander zu halten. Der Dolby Pro Logic Downmix ist deutlich zu erkennen, da die Bewegung der Bahn sich hier sehr stark auf die Stereomitte konzentriert. Die Bahngeräusche, die links und rechts in den Lautsprechern wahrzunehmen sind, stammen aus den Signalen des rechten und linken Audiokanals der 3/2-Stereo-Mischung. Durch Mischung mit der Monosumme der Surroundkanäle wirkt die Bahn so, als ob sie in der gesamten Stereobreite auf den Hörer zufährt. Auch die Räumlichkeit ist im ITU-Downmix ausgeprägter wahrzunehmen.

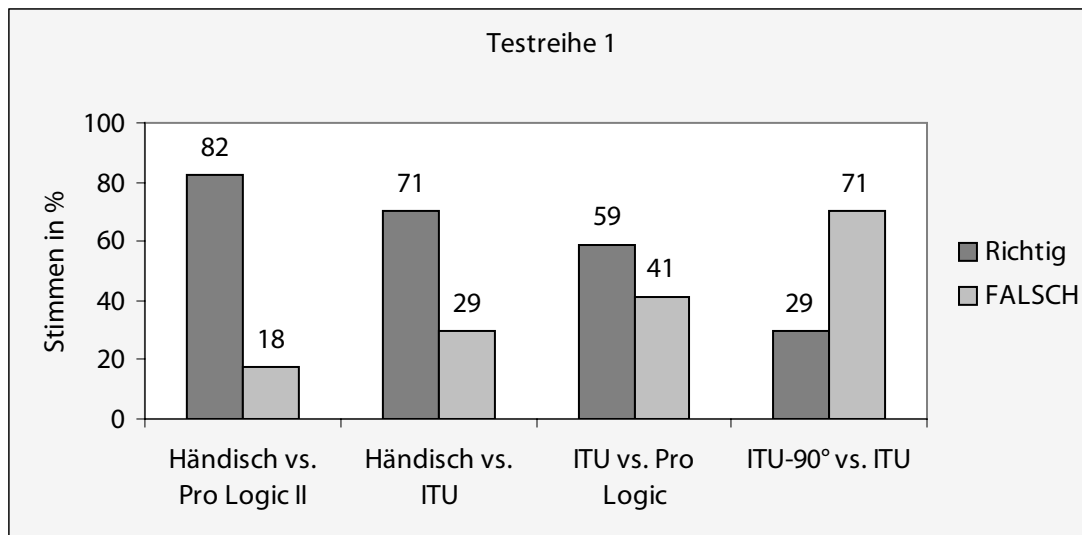


Diagramm 5-1 : Auswertung der Testreihe 1

Im Vergleich (d) wurden für die Testreihe 1 die Erwartungen erfüllt. Die beiden Hörbeispiele nach dem Prinzip des ITU-Downmix unterscheiden sich lediglich durch die bei ITU-90° Downmix in der Phase um 90° verschobenen Surroundkanäle. Da es sich bei der Bahn um eine geräuschhafte Schallquelle handelt, dürfte es sehr schwer bis unmöglich sein, hier wirklich Unterschiede wahrzunehmen. Es lagen 71 % der Probanden falsch; somit kann hier davon ausgegangen werden, dass ein Unterschied zwischen den beiden Hörbeispielen nicht wahrgenommen wurde. Das Ergebnis zeigt außerdem, dass die 90° Phasenverschiebung nicht zwingend wahrgenommen wird. Daher lässt sich auch für den Versuch (a) das insgesamt schlechte Klangbild der Dolby Pro Logic II Version nicht mit der 90° Phasenverschiebung erklären.

Testreihe 2 zeigt für das Beispiel (a), den Vergleich zwischen der händischen und der Dolby Pro Logic Version, ein noch eindeutigeres Bild als die Testreihe 1. Die Mehrzahl der Probanden waren sich einig, lediglich einer der Probanden konnte das Referenzbeispiel nicht richtig zuordnen. Die Klangverfärbungen des Dolby Pro Logic II Hörbeispiels sind auch hier deutlich wahrzunehmen, ebenso wie die geringe räumliche Tiefe. Leider ist außerdem am Ende des Audiobeispiels auf dem Schrei der Mutter in der Dolby Pro Logic II Version eine Verzerrung durch Übersteuern wahrnehmbar. Da die Unterscheidungen im Gesamtklang aber gut wahrzunehmen sind und auch in der Testreihe 1 beide Hörbeispiele gut unterschieden werden konnten, wird diese Verzerrung nicht die hauptsächliche Ursache für das eindeutige Ergebnis sein.

Der Dolby Pro Logic II Downmix der Bahnfahrt vorne links nach hinten links funktioniert prinzipiell; die Bahn scheint auf dem linken Audiokanal auf den Hörer zuzufahren. Allerdings kippt die Abbildung zwischenzeitlich auch auf die rechte Seite. Die Ursache ist wie in Testreihe 1 (a) die Downmix-Matrix der Dolby Pro Logic II Kodierung.

Vergleich (b) der Testreihe 2 zeigt, dass der Trend von Testreihe 1 beim Vergleich von ITU und händischer Version fortgesetzt wird. Nur noch 59 % der Probanden konnten das Referenzhörbeispiel richtig zuordnen. Bis auf die größere räumliche Tiefe und geringe Dynamikschwankungen beim Downmix unterscheiden sich die beiden Hörbeispiele nicht wesentlich. Die Kanaltrennung ist beim ITU-Downmix hervorragend, so dass der Downmix gut gelingt: Die Bahn fährt nur auf der linken Seite auf den Hörer zu.

Auch für die zweite Testreihe war das Ergebnis des Teils (c) überraschend. Die Probanden konnten im Schnitt keinen eindeutigen Unterschied zwischen den beiden Hörbeispielen ausmachen. Lediglich 29 % konnten das Referenzbeispiel richtig zuordnen.

Am deutlichsten ist der Unterschied zwischen dem Dolby Pro Logic und dem ITU-Downmix an der Bewegungsrichtung der Bahn zu erkennen. Beim ITU-Downmix ist die Bahn nur auf dem linken Audiokanal zu lokalisieren. Durch die Monosumme der Surroundkanäle beim Dolby Pro Logic Mixdown fährt die Bahn hingegen von der linken Seite in die Mitte. Hier kommt es dann, wie auch schon bei der Dolby Pro Logic II Mischung, zu einem instabilen Klangbild, da sich die Informationen des Signals aus der Monosumme mit den Bahngeräuschen aus dem linken und rechten Audiokanal teilweise addieren.

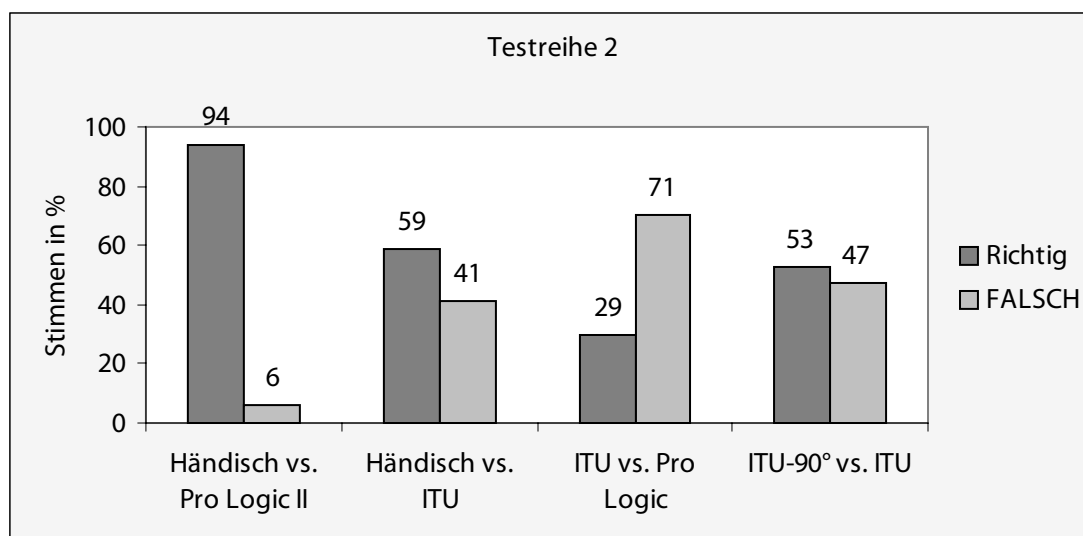


Diagramm 5-2 : Auswertung der Testreihe 2

Vergleich (d) der Testreihe 2 zeigt, dass die Vermutung, dass die 90° Phasendrehung bei geräuschhaften Audiosignalen nicht wahrzunehmen ist, richtig zu sein scheint. Zwar konnten mehr Probanden eine richtige Zuordnung treffen als in der Testreihe 1, aber bei einer Trefferquote von 51 % kann man nicht von einem klar erkennbaren Unterschied ausgehen.

Betrachtet man die relativen Häufigkeiten der Testreihen 1 und 2 gemeinsam für die einzelnen Kombinationen der Hörbeispiele, zeigt sich, dass die bisher gemachten Aussagen auch für das Gesamtbild zutreffend sind.

Der Unterschied zwischen Dolby Pro Logic II Downmix und händischer Version ist insgesamt am besten erkannt worden. Auch zwischen ITU-Downmix und händischer Version lässt sich ein Unterschied feststellen, auch wenn dieser nicht so signifikant zu sein scheint, wie zwischen Dolby Pro Logic II Downmix und händischer Version.

Die große Überraschung ist, dass der Unterschied zwischen Dolby Pro Logic Downmix und ITU-Version von den Probanden anscheinend nicht deutlich erkannt wurde. Auch eine genauere Überprüfung der Aussagen in Abhängigkeit von der Sitzposition zeigt, dass die Anzahl der richtigen und falschen Einschätzungen auf den Sitzplätzen in der Mitte und an den Seiten gleichverteilt ist.

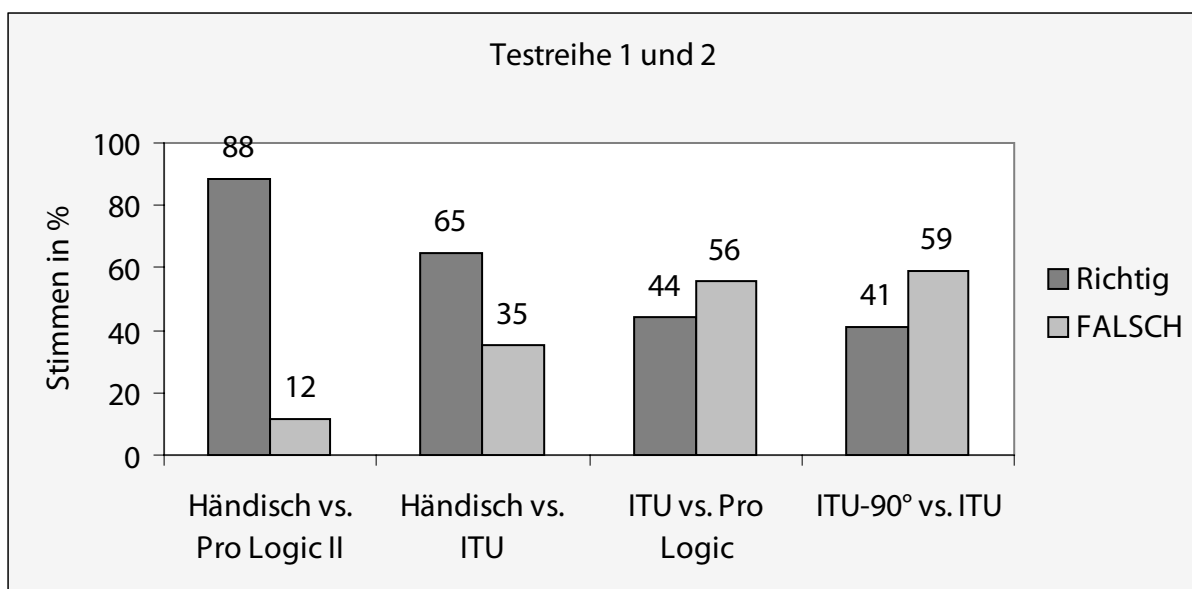


Diagramm 5-3 : Auswertung der Testreihe 1 und 2

Nachvollziehbar ist das Ergebnis jedoch für den Vergleich von ITU-90° und ITU-Downmix. Die Annahme, dass der Unterschied zwischen den Hörbeispielen bei geräuschhaftem Audiomaterial nicht wahrzunehmen ist, lässt sich auch in der Gesamtbetrachtung bestätigen.

Die Zusatzfrage, in einer Prinzipskizze einer Surround-Lautsprecheranordnung durch Pfeile, die ihrer Ansicht nach ursprüngliche Bewegungsrichtung der Phantomschallquellen einzuzeichnen, wurde von sieben Probanden beantwortet.

Für die Testreihe 1 wurde die Bewegungsrichtung von den meisten Probanden richtig erkannt, auch wenn die Bewegung den vorderen Lautsprechern zugeordnet wurde. Auch die seitliche Bewegung in Testreihe 2 konnten alle Probanden richtig zuordnen, wobei hier die Bewegungsrichtung meistens invertiert vermutet wurde. Für den Downmix von Bewegungen kann man zumindest festhalten, dass Grundzüge des dramaturgischen Konzepts richtig „übersetzt“ werden.

Die Auswertungen befinden sich im Anhang dieser Arbeit.

5.3.2 Testreihe 3

Da es sich bei der Skala zur Bewertung um eine Ordinalskala⁴⁷ handelt, darf zur Bewertung nicht das arithmetische Mittel herangezogen werden, sondern die Interpretation muss aufgrund der gegebenen Häufigkeitsverteilung erfolgen (vgl. Mayer, S. 52).

Die einzelnen Bewertungsbögen wurden manuell erfasst, nach Sitzposition geordnet, und für jeden Parameter ausgezählt, um festzustellen ob hier signifikante Unterschiede in der Stimmverteilung zu bemerken waren. Dieses war nicht der Fall. Hatten die Probanden bei einem bestimmten Parameter keine Bewertung abgegeben, wurde der Parameter unter der Kategorie „keine Angabe“ gesondert erfasst.

Nachdem für jeden der Parameter die absolute Häufigkeit ermittelt war, wurden für die Interpretation der Daten diejenigen absoluten Häufigkeiten, die kleiner als 3 waren, von der Bewertung ausgenommen. Auch die Anzahl der von der Bewertung ausgenommenen Stimmen wurde gesondert erfasst.

Von diesen absoluten Häufigkeiten wurden die relativen Häufigkeiten in Prozent bezogen auf ihre Gesamtsumme ermittelt, sowie die relativen Häufigkeiten in Prozent der nicht in die Bewertung eingeflossenen Stimmen („Keine Angabe“ und „absolute Häufigkeit kleiner als 3“), bezogen auf die Gesamtzahl der abgegebenen Stimmen. Diese Werte sollen als Grundlage der Interpretation des Hörvergleichs benutzt werden.

⁴⁷ „Die Merkmalsausprägungen bringen neben einer Verschiedenartigkeit eine natürliche Rangordnung zum Ausdruck. [...] Ordinalskalierte Merkmale sind z.B. Prüfungsnoten, Güteklassen bei Lebensmitteln und Rangplätze einer Fußballliga“ (Mayer, S.25).

Betrachtet man nun für jeden Parameter die prozentualen Anteile der Benotung für die einzelnen Hörbeispiele, muss leider festgestellt werden, dass diese für eine Aussage über die Parameter nicht eindeutig sind. Die relativen Häufigkeiten sind für fast alle Hörbeispiele und Parameter über die Ordnungszahlen 2, 3 und 4 gleich verteilt, so dass keinem der Hörbeispiele für einen bestimmten Parameter eine Tendenz zuzuschreiben wäre. Bei denjenigen Hörbeispielen hingegen, bei denen man einen Trend zu einer bestimmten Bewertung annehmen könnte, ist das Verhältnis von Stimmen die von der Bewertung ausgenommen wurden, zu denen, die zum Ergebnis beigetragen haben zu gering, um eine gute Aussage zu machen. Die Testreihe 3 soll daher nicht für weitere Aussagen über die Qualität der einzelnen Hörbeispiele, bzw. Downmix-Methoden genutzt werden.

Dieses Ergebnis ist wohl durch die zu große Anzahl der Parameter entstanden, durch welche die Probanden den einzelnen Parametern nicht genug Aufmerksamkeit zukommen lassen konnten. Für ein aussagekräftiges Ergebnis sollte bei Wiederholung eines Hörtest die Anzahl der Parameter auf einzelne, ausgewählte Hauptmerkmale und Parameter beschränkt werden. Eine Tabelle mit den relativen Häufigkeiten der Benotungen der Hörbeispiele, bezogen auf die unterschiedlichen Parameter, befindet sich im Anhang.

5.3.3 Testreihe 4

In der Testreihe 4 sollte herausgefunden werden, ob ein Unterschied zwischen einer diskreten 3/2-Stereo-Version und einer diskreten 3/2-Stereo-Version mit 90° Phasenverschiebung der Surroundkanäle, wie sie bei der Wiedergabe eines Dolby Pro Logic kompatiblen Dolby Digital Audiosignals entstehen kann, wahrgenommen wird.

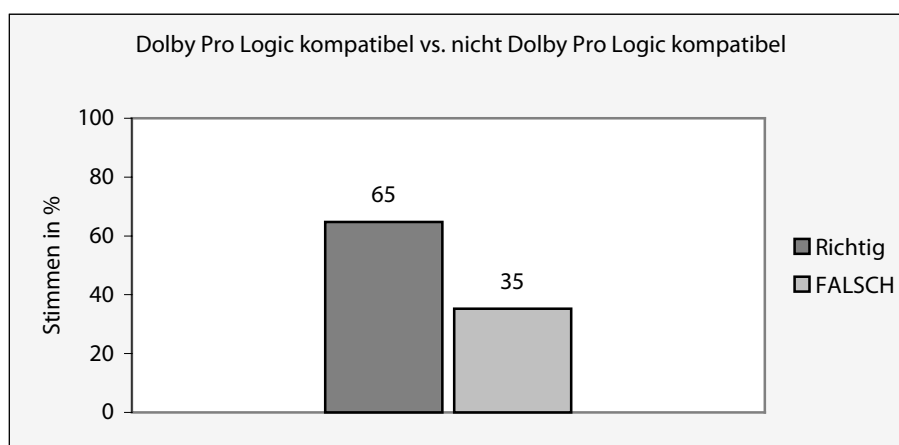


Diagramm 5-4 : Auswertung der Testreihe 4

Das Testergebnis zeigt, das anscheinend 65 % der Probanden hier Unterschiede wahrgenommen haben. Ich bin jedoch der Ansicht, dass bei geräuschhaften Audiosignalen, wie bei den hier verwendeten U-Bahngeräuschen, eine Unterscheidung nicht möglich sein sollte. Allerdings kann es sein, dass durch die geringere Korrelation der Audiosignale beim Dolby Pro Logic kompatiblen 3/2-Stereosignal die seitlichen Phantomschallquellen instabiler wirken und die Probanden diesen Effekt unterschwellig wahrgenommen haben. Im persönlichen Gespräch mit den Probanden nach Beendigung des Hörvergleichs wurde oft geäußert, dass eines der beiden Hörbeispiele im Vergleich als „voller“ empfunden wurde. Dieses könnte die Vermutung bestätigen. Die Auswertung befindet sich im Anhang dieser Arbeit.

5.4 Gesamtbetrachtung

Das Ergebnis des Hörvergleichs wirft einige Fragen auf, die hier diskutiert werden sollen. Zunächst ist die Auffälligkeit, dass der Unterschied zwischen der händischen 2/0-Stereo-Version und dem Dolby Pro Logic II Downmix in dieser Deutlichkeit ausfällt. Florian Rathgeber hatte in einer ähnlichen Arbeit festgestellt, dass die Klangqualität des Dolby Pro Logic II Downmix einer händischen Version nicht wesentlich nachsteht; und das ZDF nutzt die Möglichkeiten des automatischen Dolby Pro Logic II Downmix mit Erfolg für seine Live-Sendungen. Die Erwartung war daher, dass der Unterschied nicht in dieser Deutlichkeit wahrnehmbar sein sollte.

Auffällig sind bei dem von der Firma Dolby hergestellten Dolby Pro Logic II Hörbeispiel die Klangfärbungen des Audiomaterials. Durch die teilweise zu hoch ausgesteuerte Mischung, sind Verzerrungen wahrzunehmen. Da die Kodierung nicht optimiert werden konnte und nur mit den Standardparametern durchgeführt wurde, könnte es sein, dass bei der Kodierung noch Verbesserungen für die Klangqualität möglich sind. Außerdem ist der Dolby Pro Logic II Downmix die einzige 2/0-Stereo-Version, die nicht mit der Audiosoftware NUENDO, also ohne eventuelle Zwischenwandlung, hergestellt wurde. Ob dieser Faktor eine Rolle spielt, kann zum derzeitigen Zeitpunkt nicht eindeutig festgestellt werden.

Die weniger starke Kanaltrennung dürfte für das Ergebnis keine wesentliche Rolle gespielt haben, da für die Dolby Pro Logic kompatiblen Hörbeispiele ein den Erwartungen genau Gegenteiliges Ergebnis ermittelt wurde. Die eigentlich gut wahrnehmbaren Klangverluste des Dolby Pro Logic Downmix wurden von den Hörern so gut wie nicht wahrgenommen.

Allerdings stellt sich die Frage, ob ein Dolby Pro Logic Downmix professionellen Ansprüchen genügt, und ob dieses Format generell noch als zeitgemäß bezeichnet werden kann.

Im Bezug auf die Klangqualität scheint der ITU-Downmix einen guten Kompromiss darzustellen. Es wurde zwar ein Unterscheid zur händischen Version wahrgenommen, jedoch ist dieser Unterschied nicht so eindeutig wie zum Dolby Pro Logic II Downmix. Für Audiomaterial das stark korrelierte Signale in den einzelnen Kanälen enthält und auf eine gute Kanaltrennung nach dem Downmix angewiesen ist, kann der ITU-Downmix gut verwendet werden. Um die technischen Fehler zu verhindern, wäre auch der ITU-90° Downmix eine Alternative, da ein Unterschied zwischen den beiden Downmix-Varianten anscheinend nicht hörbar ist. Allerdings muss noch genauer überprüft werden, ob sich die Klangqualität bei der diskreten Wiedergabe nicht wesentlich verschlechtert.

Der Hörvergleich hat außerdem am Beispiel des Kurzhörspiels gezeigt, dass der Einsatz von Äquivalenzstereophonie zur Klanggestaltung gut funktioniert. Die gewünschten Bewegungen sind sowohl in der 3/2-Stereo-Wiedergabe als auch nach dem Downmix gut wahrzunehmen.

6 Gesamtreflexion und Ausblick

Die theoretischen und praktischen Untersuchungen haben gezeigt, dass die Downmix-Kompatibilität von Mehrkanalaudio-Produktionen bei bewegten Signalen unter gewissen Voraussetzungen ohne stark wahrnehmbare Klangverluste gegeben ist. Je nach Downmix-Variante ergeben sich unterschiedliche Vor- und Nachteile.

Der Downmix hat immer eine Reduktion der Audiokanäle der Originalmischung auf zwei Audiokanäle zur Folge. Für die Qualität des Downmix ist entscheidend, dass es durch das Zusammenfassen von mehreren Signalen, nicht zu Auslöschungen oder Übersteuern kommt. Um die Downmix-Kompatibilität von Mehrkanalaudio-Produktionen zu unterstützen, sollten bei der Erstellung des 3/2-Stereosignals die Auswirkungen des Downmix immer mitbedacht werden. Dies betrifft sowohl die spätere Position der Audiosignale im Stereobild sowie technische Fehler.

Das Phänomen der Phantomschallquelle ermöglicht dem Hörer Bewegungen von Klangereignissen im 3/2-Stereobild wahrzunehmen. Um den Eindruck von stabilen Phantomschallquellen zu schaffen, müssen die Signale der beteiligten Lautsprecher eine hohe Korrelation aufweisen. Da die 3/2-Stereo-Standard-Aufstellung gerade im Bezug auf

seitliche und hintere Phantomschallquellen nicht ideal ist, sind grundlegende Kenntnisse über die menschliche Wahrnehmung und das räumliche Hören notwendig, um Klangeindrücke zu schaffen, die vom Hörer nachvollzogen werden können. Der Downmix bedeutet für Klangeindrücke, die mit Hilfe der Surround-Lautsprecher erzeugt wurden, wie dem Eindruck von Räumlichkeit oder bewegte Schallquellen, einen Übergang von einer natürlichen Wahrnehmung zu einer simulierten.

Der Downmix von Bewegungen in den vorderen Lautsprechern ist in der Regel unproblematisch, da bis auf den Centerkanal, die an der Bewegung beteiligten Audiokanäle erhalten bleiben. Für Bewegungen seitlich oder hinter dem Hörer bedeutet der Downmix jedoch, dass die an der Bildung der Phantomschallquelle beteiligten Audiokanäle nicht mehr vorhanden sind und somit der Eindruck einer bewegten Schallquelle nur noch als Simulation möglich ist. Der Eindruck einer bewegten Schallquelle kann auch nach dem Downmix erhalten bleiben, wenn das Originalsignal zusätzliche Informationen enthält, die über die Art der Bewegung Auskunft geben. Dieses kann z.B. ein gewisser Grad an Räumlichkeit sein.

Da der Eindruck von Räumlichkeit das Vorhandensein von unkorrelierten Signalen voraussetzt, müssen bei der Produktion Kompromisse getroffen werden. Um downmix-kompatible Audiosignale zu erzeugen, die ein ausgeglichenes Verhältnis an korrelierten und nicht korrelierten Signalen aufweisen, eignet sich die Äquivalenzstereophonie besser als reine Laufzeit- oder Pegeldifferenzen.

Für die Downmix-Varianten Dolby Pro Logic/ Pro Logic II ist die große Korrelation der Audiosignale allerdings nachteilig, da Dolby voraussetzt, dass die Audiosignale in den Surround-Kanälen unkorreliert sind. Beim Dolby Pro Logic II Downmix ergibt sich durch eine große Korrelation zwischen den vorderen Audiokanälen und den Surround-Kanälen bzw. zwischen den Surround-Kanälen ein undifferenziertes Klangbild nach dem Downmix. Audiosignale, die ursprünglich nur der linken oder rechten Seite zugeordnet waren, finden sich nach dem Downmix auch in der entsprechend entgegengesetzten Abbildungsrichtung wieder.

Der Dolby Pro Logic Downmix führt durch die Monosumme der Surroundkanäle zu erheblichen Klang- und Abbildungsverlusten. Die Simulation von Klangeindrücken nach dem Downmix ist nur noch eingeschränkt möglich, da sich die Audiosignale aus den Surround-Kanälen lediglich als Monosumme in der Stereomitte wiederfinden.

Für einen Downmix nach ITU-R 775 ist die Korrelation der Signale unproblematisch, da hier der linke und rechte Surround-Kanal nur dem linken und rechten Audiokanal zugemischt werden und es keine „vertauschten Seiten“ gibt.

Der Hörvergleich hat gezeigt, dass der ITU-Downmix als Alternative zur händischen Version akzeptiert wird. Technische Fehler für Signale zwischen linkem Surroundkanal und linkem Audiokanal bzw. rechtem Surroundkanal und rechtem Audiokanal können mit der Variante des ITU-90° Downmix vermieden werden. Allerdings muss kontrolliert werden, ob sich durch die 90° Phasenverschiebung Nachteile für die diskrete 3/2-Stereo-Wiedergabe ergeben.

Es stellt sich die Frage welchen Vorteil die Beschäftigung mit alternativen Downmix-Varianten wie Dolby Pro Logic II oder der Forderung nach ganz neuen Algorithmen hat, oder ob man sich in Zukunft darauf konzentrieren sollte, das Audiomaterial so anzupassen, das beispielsweise der ITU-Downmix in jedem Fall zu einem guten Ergebnis führt. Mit Hilfe von Dolby Metadaten sollte es möglich sein, das Audiomaterial so anzupassen, dass die Wiedergabequalität professionellen Ansprüchen genügt. Da mit Einführung von HDTV der automatische Downmix eines Dolby Digital Signals im Decoder ohne Probleme möglich sein wird, sollte man diese Alternative ernsthaft in Betracht ziehen, anstatt auf die Übertragung von zwei unabhängigen Tonspuren zu setzen. Die Möglichkeiten mit den Dolby Matrix-Verfahren eine Decodierung des Downmix vorzunehmen und so eine Mehrkanal Wiedergabe zu ermöglichen, sind meiner Ansicht nach kein Argument für einen Einsatz dieser Formate. Beide Audioformate werden meiner Ansicht nach in Zukunft durch die Möglichkeit der digitalen Distribution und den Einsatz des Dolby Digital Formats eine immer geringere Rolle spielen.

7 Literatur

- **Adam, Carsten:** Artefakte im Multichannel- Audio in der Dolby Produktionskette beim ZDF. Düsseldorf: Diplomarbeit. Fachhochschule Düsseldorf. Fachbereich Medien. Studiengang Ton- und Bildtechnik, 2004.
- **Birkner, Christian:** Surround. Einführung in die Mehrkanalton- Technik. Bergkirchen: PPV Presse Project Verlags GmbH, 2002.
- **Blauert, Jens:** Räumliches Hören. Stuttgart: S. Hirzel Verlag, 1974.
- **Blauert, Jens:** Räumliches Hören. Nachschrift. Neue Ergebnisse und Trends seit 1972. Stuttgart: S. Hirzel Verlag, 1985.
- **Carroll, Tim:** What Is Dolby E? TV Technology.com, 2003.
URL.: http://www.tvtechnology.com/features/audio_notes/f_tc_audio_notes-06.25.03.shtml
(aufgerufen am 10.10.2005)
- **Dettwiler, Daniel:** Untersuchung zur Positionierung von Schallereignissen im Phantomfeld der Fünfkanal- Stereophonie bei der Mischung von U- Musik. Basel, Elektronisches Studio der Musikhochschule Basel (2001)
URL.: <http://www.esbasel.ch/Studierende/studentenarbeiten/Sourround/SouPos.PDF> (aufgerufen am 17.11.2005)
- **Dickreiter, Michael:** Handbuch der Tonstudioteknik. 6. Auflage. München: K.G. Sauer Verlag KG, 1997.
- **Gernemann, Andreas:** Kohärenz und Korrelation in der Tonstudioteknik
URL.: <http://www.uni-koeln.de/phil-fak/muwi/ag/tec/kohkor.pdf> (aufgerufen am 25.11.2005)

- **Griesinger, David:** Multichannel Matrix Surround Decoders for Two-Eared Listeners. URL.: <http://world.std.com/~griesngr/sur.pdf> (aufgerufen am 24.10.2005)
- **Griesinger, David:** A note to Eberhard Sengpiel concerning Theile's experiments on the correlation of reverberation.
URL.: http://world.std.com/~griesngr/note_to_Sengpiel.txt (aufgerufen am 07.11.05)
- **Griesinger, David:** Surround From Stereo. Power Point Präsentation. 2002.
URL.: http://world.std.com/~griesngr/surround_from_stereo2.ppt (aufgerufen am 24.10.2005)
- **Hilton, Kevin:** Surround TV. Erste Europäische 5.1 Liveübertragung. In: Audio Media. (Cambridgeshire)Heft Nr.12 (2003), Seite 34 – 37.
- **Katz, Bob:** Masterin Audio. The Art and Science. Oxford: Focal Press, 2002.
- **Magarelli, Romolo; Strachan, David:** Integrated Solutions for Embedded Dolby E and AC-3. URL.: http://www.roscor.com/resources_whitepapers/Dolby_Integrated_Solutions.doc (aufgerufen am 10.10. 2005)
- **Mayer, Horst:** Beschreibende Statistik. 4.Auflage. München: Carl Hanser Verlag, 2006.
- **Mitchel, Doug:** Surround Setup Szenarios. In: Audio Media. (Cambridgeshire) Heft Nr.8 (2002), Seite 44- 45.
- **Pohlmann, Ken C.:** Principles of Digital Audio. 4. Auflage. New York: McGraw- Hill, 2000.

- **Rathgeber, Florian:** Down- Mixing Compatibility within Multi- Channel Audio Productions – Lt/Rt Signals -. Diplomarbeit im Studiengang Fernsehtechnik und elektronische Medien, Fachhochschule Wiesbaden, Fachbereich Elektrotechnik. Wiesbaden, 2004.
- **[P&E] The Recording Academy’s Producers & Engineers Wing:** Recommendations For Surround Sound Production. Santa Monica, 2004.
URL.: http://www.grammy.com/pe_wing/5_1_Rec.pdf (aufgerufen am 19.09.2005)
- **Ruprecht, Werner:** Signale und Übertragungssysteme. Berlin: Springer Verlag, 1993.
- **Sengpiel, Eberhard:** Vorlesungsunterlagen.
URL.: <http://www.sengpielaudio.com> (aufgerufen am 19.09.2005)
- **Steinke, Gerhard:** Wie viel Kanäle/Signale braucht der Mensch?. Plädoyer für die Standard-3/2-Stereo-Hierarchie und ihre Optimierung im Heim. Hannover: Internet-Beitrag zur 21. Tonmeistertagung des VDT, 2003.
URL.: www.mdr.de/DL/939974.doc (aufgerufen am 19.09.2005)
- **Steveaux, Andreas:** Integrationsmöglichkeit der Audio-Metadatensteuerung für einen Mehrkanalhörfunkdienst und Kompatibilitätsuntersuchungen an Empfangsgeräten bei Ausstrahlung nach dem DVB-S Standart. Diplomarbeit. Köln: Fachhochschule Köln, Institut für Nachrichtentechnik, 2005.
- **Theile, Günter:** Grundlagen und Praxis der Surround- Technik. Leipzig: MDR - 64. VDT- Seminar 24.-26. Oktober, 2003.

- **Theile, Günter:** On the Naturalness of Two-Channel Stereo Sound. In: Journal of the AES, Vol.39; 1991.
URL.: http://www.irt.de/wittek/hauptMikrophon/theile/Theile_Naturelness_JAES1991.PDF
(aufgerufen am 08.11.2005)
- **Theile, Günter:** Multichannel Natural Music Recordings Based On Psychoacoustic Principles. München: 2001. 1. und 2. Version.
URL.:http://www.irt.de/wittek/hauptMikrophon/theile/Multich_Recording_30.Oct.2001_.PDF
(aufgerufen am 04.11.2005)
- **Veit, Ivar:** Technische Akustik. 5. Auflage. Würzburg: Vogel, 1996.
- **Watkinson, John:** An Introduction to Digital Audio. Oxford: Focal Press, 1994.
- **Werner, Martin:** Nachrichtentechnik. Eine Einführung für alle Studiengänge. 4. Auflage. Wiesbaden : Vieweg, 2003.
- **Wittek, Helmut :** Zur Problematik des automatischen Downmix von Mehrkanal auf Zweikanal. Downmix Referat. Nürnberg, 2005.
URL.: http://hauptmikrophon.de/Downmix_SRT_Nuernberg_0505.pdf (aufgerufen am 25.10.2005)
- **[A/52B]** - Digital Audio Compression (AC-3) (E-AC3) Standard, Rev. B (2005). ATSC, Washington, D.C. 20006.
URL.: http://www.atsc.org/standards/a_52b.pdf (aufgerufen am 28.09.2005)

Dolby Laboratories, Inc Aufsätze

Alle hier aufgeführten Aufsätze befinden sich auf der Internetpräsenz der Firma Dolby Laboratories, Inc unter: **http://www.Dolby.com/resources/tech_library/index.cfm**

- **[1]** Mixing Information for Dolby Pro Logic II. San Francisco: Dolby Laboratories, Inc, 2005.
URL.: http://www.Dolby.com/assets/pdf/tech_library/214_Mixing%20with%20Dolby%20Pro%20Logic%20II%20Technology.pdf (aufgerufen am 19.09.2005)
- **[2]** Dolby Surround Pro Logic II Decoder Principles of Operation.
URL.: http://www.Dolby.com/assets/pdf/tech_library/209_Dolby_Surround_Pro_Logic_II_Decoder_Principles_of_Operation.pdf (aufgerufen am 19.09.2005)
- **[3]** Dolby DP563- Dolby Surround and Pro Logic II Encoder User's Manual. Issue 3. San Francisco: Dolby Laboratories, Inc , 2003.
URL.: http://www.Dolby.com/assets/pdf/tech_library/148_563_2.Manual.pdf (aufgerufen am 19.09.2005)
- **[4]** Dolby Surround Mixing Manual. Issue 2. San Francisco: Dolby Laboratories, Inc, 2005.
URL.: http://www.Dolby.com/assets/pdf/tech_library/44_SurroundMixing.pdf (aufgerufen am 19.09.2005)
- **[5]** Dolby Digital Professional Encoding Guidelines. Issue 1. San Francisco: Dolby Laboratories, Inc, 2000.
URL.: http://www.Dolby.com/assets/pdf/tech_library/46_DDEncodingGuidelines.pdf (aufgerufen am 19.09.2005)
- **[6]** Dolby Metadata Guide. Issue 2. San Francisco: Dolby Laboratories, Inc, 2003.
URL.: http://www.Dolby.com/assets/pdf/tech_library/18_Metadata.Guide.pdf (aufgerufen am 19.09.2005)

- **[7]** Todd, Craig, u.a.; AC-3: Flexible Coding for Audio Transmission and Storage. San Francisco: Dolby Laboratories, Inc, 1994.
URL.: http://www.Dolby.com/assets/pdf/tech_library/37_ac3-flex.pdf (aufgerufen am 26.09.2005)
- **[8]** Davis, Mark F.: The AC-3 Multichannel Coder. San Francisco: Dolby Laboratories, Inc, 1993.
URL.: http://www.Dolby.com/assets/pdf/tech_library/203_ac3_multichannel_decoder.pdf (aufgerufen am 26.09.2005)
- **[9]** Dolby E, Multichannel Audio Coding for DTV Production and Distribution. San Francisco: Dolby Laboratories, Inc, 1993.
URL.: http://www.Dolby.com/assets/pdf/tech_library/49_Dolby_E_Innovation.pdf (aufgerufen am 10.10.2005)

EBU Tech 3000 series

Alle hier aufgeführten Aufsätze befinden sich auf der Internetpräsenz der European Broadcast Union unter:

http://www.ebu.ch/en/technical/publications/tech3000_series/index.php

(aufgerufen am 11.12.2005)

- **[1]** EBU Tech. 3276: Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic. 2te Auflage (1998).
- **[2]** Tech 3276-E: Listening conditions for the assessment of sound programme material. Supplement 1. Multichannel sound. (2004)
- **[3]** Tech. 3286-E: Assessment methods for the subjective evaluation of the quality of sound programme material – Music. (1997)

8 Anhang

| | | |
|-----------|--|----|
| 1 | Testreihe 1: Auswertung der Testbögen vom 02. und 05. Dezember 2005 in absoluten Häufigkeiten..... | 1 |
| 2 | Testreihe 2: Auswertung der Testbögen vom 02. und 05. Dezember 2005 in absoluten Häufigkeiten..... | 2 |
| 3 | Gesamtbetrachtung der Testreihe 1 und 2, bezogen auf die unterschiedlichen Audioformate. | 3 |
| 4 | Auswertung der Testreihe 3..... | 4 |
| 5 | Testreihe 4: Auswertung der Testbögen vom 02. und 05. Dezember 2005 in absoluten Häufigkeiten..... | 9 |
| 6 | Bewertungsbogen für Testreihe 1, 2 und 4..... | 10 |
| 7 | Bewertungsbogen für Testreihe 3 | 11 |
| 8 | Bewertungsskala für subjektive Einschätzungen nach EBU-Tech 3286 (1997) | 12 |
| 9 | Hörspielskript..... | 15 |
| 10 | Lageplan U-Bahnhaltestelle Hamburg-Jungfernstieg..... | 20 |

1. Testreihe 1: Auswertung der Testbögen vom 02. und 05. Dezember 2005 in absoluten Häufigkeiten

| Test 1 - 02.12. | Pro Logic II vs. händisch | | ITU vs. händisch | | ITU vs. Pro Logic | | ITU-90° vs. ITU | |
|----------------------|---------------------------|-----------|------------------|-----------|-------------------|-----------|-----------------|-----------|
| | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH |
| Gruppe 1 | 4 | | 4 | | 2 | 2 | | 4 |
| Gruppe 2 | 1 | 1 | 1 | 1 | 2 | | 2 | |
| Gruppe 3 | 3 | | 3 | | 2 | 1 | | 3 |
| Gesamt | 8 | 1 | 8 | 1 | 6 | 3 | 2 | 7 |
| | | | | | | | | |
| Test 1 -05.12. | Pro Logic II vs. händisch | | ITU vs. händisch | | ITU vs. Pro Logic | | ITU-90° vs. ITU | |
| | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH |
| Gruppe 1 | 2 | | 1 | 1 | 1 | 1 | 2 | |
| Gruppe 2 | 2 | | 1 | 1 | 1 | 1 | 1 | 1 |
| Gruppe 3 | | 2 | 1 | 1 | 1 | 1 | | 2 |
| Gruppe 4 | 2 | | 1 | 1 | 1 | 1 | | 2 |
| Gesamt | 6 | 2 | 4 | 4 | 4 | 4 | 3 | 5 |
| | | | | | | | | |
| Gesamt Gr.1+2 | 14 | 3 | 12 | 5 | 10 | 7 | 5 | 12 |
| | | | | | | | | |
| Gesamt in % | 82 | 18 | 71 | 29 | 59 | 41 | 29 | 71 |

2. Testreihe 2: Auswertung der Testbögen vom 02. und 05. Dezember 2005 in absoluten Häufigkeiten

| Test 2 - 02.12 | Pro Logic II vs. händisch | | ITU vs. händisch | | ITU vs. Pro Logic | | ITU-90° vs. ITU | |
|--------------------|---------------------------|----------|------------------|-----------|-------------------|-----------|-----------------|-----------|
| | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH |
| Gruppe 1 | 4 | | 4 | | | 4 | 3 | 1 |
| Gruppe 2 | 2 | | 1 | 1 | | 2 | 2 | |
| Gruppe 3 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| Gesamt | 8 | 1 | 6 | 3 | 2 | 7 | 6 | 3 |
| | | | | | | | | |
| Test 2 - 05.12. | Pro Logic II vs. händisch | | ITU vs. händisch | | ITU vs. Pro Logic | | ITU-90° vs. ITU | |
| | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH |
| Gruppe 1 | 2 | | 2 | | 2 | | 1 | 1 |
| Gruppe 2 | 2 | | 1 | 1 | | 2 | 2 | |
| Gruppe 3 | 2 | | | 2 | | 2 | | 2 |
| Gruppe 4 | 2 | | 1 | 1 | 1 | 1 | | 2 |
| Gesamt | 8 | | 4 | 4 | 3 | 5 | 3 | 5 |
| | | | | | | | | |
| Gesamt 1+2 | 16 | 1 | 10 | 7 | 5 | 12 | 9 | 8 |
| Gesamt in % | 94 | 6 | 59 | 41 | 29 | 71 | 53 | 47 |

3. Gesamtbetrachtung der Testreihe 1 und 2, bezogen auf die unterschiedlichen Audioformate.

| Test am 02.12 | Pro Logic II vs. händisch | | ITU vs. händisch | | ITU vs. Pro Logic | | ITU-90° vs. ITU | |
|----------------------|---------------------------|-----------|------------------|-----------|-------------------|-----------|-----------------|-----------|
| | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH |
| Gruppe 1 | 8 | | 8 | | 2 | 6 | 3 | 5 |
| Gruppe 2 | 3 | 1 | 2 | 2 | 2 | 2 | 4 | |
| Gruppe 3 | 5 | 1 | 4 | 2 | 4 | 2 | 1 | 5 |
| | | | | | | | | |
| Gesamt | 16 | 2 | 14 | 4 | 8 | 10 | 8 | 10 |
| | | | | | | | | |
| Test am 05.12 | Pro Logic II vs. händisch | | ITU vs. händisch | | ITU vs. Pro Logic | | ITU-90° vs. ITU | |
| | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH | Richtig | FALSCH |
| Gruppe 1 | 4 | | 3 | 1 | 3 | 1 | 3 | 1 |
| Gruppe 2 | 4 | | 2 | 2 | 1 | 3 | 3 | 1 |
| Gruppe 3 | 2 | 2 | 1 | 3 | 1 | 3 | | 4 |
| Gruppe 4 | 4 | | 2 | 2 | 2 | 2 | | 4 |
| Gesamt | 14 | 2 | 8 | 8 | 7 | 9 | 6 | 10 |
| | | | | | | | | |
| Gesamt Gr.1+2 | 30 | 4 | 22 | 12 | 15 | 19 | 14 | 20 |
| | | | | | | | | |
| Gesamt in % | 88 | 12 | 65 | 35 | 44 | 56 | 41 | 59 |

4. Auswertung der Testreihe 3

| Parameter | Audiobeispiel | Note 1 in % | Note 2 in % | Note 3 in % | Note 4 in % | Note 5 in % | Note 6 in % | Ausgemusterte % der gesamten Stimmen |
|---------------------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------------------------------|
| Homogenität des Raumes | ITU | | 31 | 69 | | | | 24 |
| | ITU-90° | | 46 | 54 | | | | 24 |
| | Dolby Pro Logic | | 50 | 31 | 19 | | | 6 |
| | Dolby Pro Logic II | | 38 | 38 | 23 | | | 24 |
| | Händisch | | 57 | 43 | | | | 18 |
| Nachhall | ITU | | 38 | 25 | 19 | 19 | | 6 |
| | ITU-90° | | | 100 | | | | 41 |
| | Dolby Pro Logic | | 31 | 44 | 25 | | | 6 |
| | Dolby Pro Logic II | | 31 | 38 | 31 | | | 24 |
| | Händisch | | 50 | 31 | 19 | | | 6 |
| Akustische Ausgewogenheit | ITU | | 33 | 67 | | | | 29 |
| | ITU-90° | | 38 | 62 | | | | 24 |
| | Dolby Pro Logic | | 33 | 40 | 27 | | | 12 |
| | Dolby Pro Logic II | | 50 | 25 | 25 | | | 6 |
| | Händisch | | 29 | 50 | 21 | | | 18 |

| Parameter | Audiobeispiel | Note 1 in % | Note 2 in % | Note 3 in % | Note 4 in % | Note 5 in % | Note 6 in % | Ausgemusterte % der gesamten Stimmen |
|--------------------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------------------------------|
| Ersichtliche Raumgröße | ITU | | 20 | 40 | 20 | 20 | | 12 |
| | ITU-90° | | 25 | 75 | | | | 29 |
| | Dolby Pro Logic | | | 38 | 62 | | | 24 |
| | Dolby Pro Logic II | | 23 | 38 | 38 | | | 24 |
| | Händisch | | 19 | 44 | 38 | | | 6 |
| Tiefenstaffelung | ITU | | 36 | 43 | 21 | | | 18 |
| | ITU-90° | | 31 | 31 | 38 | | | 6 |
| | Dolby Pro Logic | | 25 | 31 | 44 | | | 6 |
| | Dolby Pro Logic II | | 38 | 25 | 38 | | | 6 |
| | Händisch | | 33 | 40 | 27 | | | 12 |
| Klangfarbe des Nachhalls | ITU | | 33 | 67 | | | | 29 |
| | ITU-90° | | 27 | 73 | | | | 35 |
| | Dolby Pro Logic | | 21 | 57 | 21 | | | 18 |
| | Dolby Pro Logic II | | 50 | 25 | 25 | | | 29 |
| | Händisch | | 29 | 50 | 21 | | | 18 |

| Parameter | Audiobeispiel | Note 1 in % | Note 2 in % | Note 3 in % | Note 4 in % | Note 5 in % | Note 6 in % | Ausgemusterte % der gesamten Stimmen |
|------------------------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------------------------------|
| Ausgewogenheit d. Richtungen | ITU | | 64 | 36 | | | | 18 |
| | ITU-90° | | 67 | 33 | | | | 29 |
| | Dolby Pro Logic | | 53 | 18 | 29 | | | |
| | Dolby Pro Logic II | | 100 | | | | | 53 |
| | Händisch | | 47 | 27 | 27 | | | 12 |
| Stabilität | ITU | | 50 | 50 | | | | 29 |
| | ITU-90° | | 45 | 55 | | | | 35 |
| | Dolby Pro Logic | | 58 | 42 | | | | 29 |
| | Dolby Pro Logic II | 27 | 36 | 36 | | | | 35 |
| | Händisch | 19 | 31 | 31 | 19 | | | 6 |
| Breite des Hörbildes | ITU | | 50 | 31 | 19 | | | 6 |
| | ITU-90° | | 43 | 36 | 21 | | | 18 |
| | Dolby Pro Logic | | 64 | 36 | | | | 18 |
| | Dolby Pro Logic II | | 69 | 31 | | | | 24 |
| | Händisch | | 100 | | | | | 29 |

| Parameter | Audiobeispiel | Note 1 in % | Note 2 in % | Note 3 in % | Note 4 in % | Note 5 in % | Note 6 in % | Ausgemusterte % der gesamten Stimmen |
|---------------------------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------------------------------|
| Lokalisationsschärfe | ITU | 24 | 41 | 18 | 18 | | | 0 |
| | ITU-90° | | 77 | 23 | | | | 24 |
| | Dolby Pro Logic | | 100 | | | | | 35 |
| | Dolby Pro Logic II | | 67 | | 33 | | | 47 |
| | Händisch | 25 | 44 | 31 | | | | 6 |
| Zuordnung Klangerzeuger – Klang | ITU | 24 | 47 | 29 | | | | 0 |
| | ITU-90° | 21 | 50 | 29 | | | | 18 |
| | Dolby Pro Logic | | 75 | 25 | | | | 6 |
| | Dolby Pro Logic II | 23 | 77 | | | | | 24 |
| | Händisch | 23 | 77 | | | | | 24 |
| Lautstärkeverhältnisse | ITU | | 50 | 50 | | | | 29 |
| | ITU-90° | | 64 | 36 | | | | 35 |
| | Dolby Pro Logic | | 53 | 27 | 20 | | | 12 |
| | Dolby Pro Logic II | | 75 | | | 25 | | 29 |
| | Händisch | | 79 | 21 | | | | 18 |

| Parameter | Audiobeispiel | Note 1 in % | Note 2 in % | Note 3 in % | Note 4 in % | Note 5 in % | Note 6 in % | Ausgemusterte % der gesamten Stimmen |
|----------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------------------------------|
| Dynamikumfang | ITU | | 50 | 50 | | | | 18 |
| | ITU-90° | | 50 | 50 | | | | 29 |
| | Dolby Pro Logic | 19 | 50 | 31 | | | | 6 |
| | Dolby Pro Logic II | | 60 | 40 | | | | 41 |
| | Händisch | 27 | 53 | 20 | | | | 12 |
| Gesamteindruck | ITU | | 47 | 53 | | | | 12 |
| | ITU-90° | | 50 | 50 | | | | 18 |
| | Dolby Pro Logic | | 38 | 44 | 19 | | | 6 |
| | Dolby Pro Logic II | | 53 | 47 | | | | 12 |
| | Händisch | | 56 | 25 | 19 | | | 6 |

Für die Werte der relativen Häufigkeiten der Ordinalzahlen gilt:

$$\text{Wert} = \text{Stimmenanzahl} \cdot \frac{100}{\text{Gesamtstimmen} - \text{"ungültige Stimmen"}}$$

Für die Werte der Ausgemusterte % der Gesamten Stimmen gilt:

$$\text{Wert} = \text{"ungültige" Stimmen} \cdot \frac{100}{\text{Gesamtstimmen}}$$

5. Testreihe 4: Auswertung der Testbögen vom 02. und 05. Dezember 2005 in absoluten Häufigkeiten

| Test 4 - 02.12 | 3/2- Stereo vs. 3/2- Stereo Pro Logic kompatibel | |
|----------------|--|----------|
| | Richtig | FALSCH |
| Gruppe 1 | 4 | |
| Gruppe 2 | 2 | |
| Gruppe 3 | 2 | 1 |
| Gesamt | 8 | 1 |

| Test 4 - 05.12. | I3/2- Stereo vs. 3/2- Stereo Pro Logic kompatibel | |
|-----------------|---|----------|
| | Richtig | FALSCH |
| Gruppe 1 | | 2 |
| Gruppe 2 | 2 | |
| Gruppe 3 | 1 | 1 |
| Gruppe 4 | | 2 |
| Gesamt | 3 | 5 |

| | | |
|------------|----|---|
| Gesamt 1+2 | 11 | 6 |
|------------|----|---|

| | | |
|-------------|----|----|
| Gesamt in % | 65 | 35 |
|-------------|----|----|

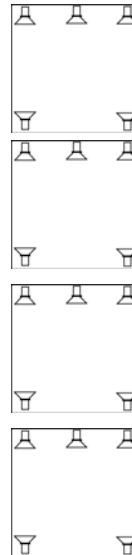
6. Bewertungsbogen für Testreihe 1, 2 und 4

Sitzplatz Nr. _____

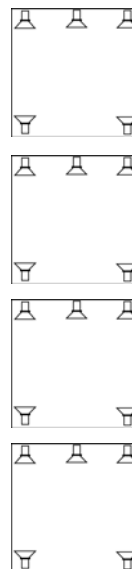
Datum: _____

Gruppe: _____

1. Hörspiel Szene 1

a) 1 2 b) 1 2 c) 1 2 d) 1 2 

2. Hörspiel Szene 2 - ABX

a) 1 2 b) 1 2 c) 1 2 d) 1 2 

4. Surround Sound

a. 1 2 b. 1 2

7. Bewertungsbogen für Testreihe 3

Bewertungsbogen für Hörbeispiel Nr. _____

Sitzplatz Nr. _____

Datum: _____

Gruppe: _____

| Hauptmerkmale | Zugeordnete Merkmale | | | | | | |
|--|---------------------------------|---|---|---|---|---|---|
| 1. Räumlicher Eindruck Das Klagereignis findet in einem eindeutig definiertem räumlichen Umfeld statt | Homogenität des Raumes | 1 | 2 | 3 | 4 | 5 | 6 |
| | Nachhall | 1 | 2 | 3 | 4 | 5 | 6 |
| | Akustische Ausgewogenheit | 1 | 2 | 3 | 4 | 5 | 6 |
| | Ersichtliche Raumgröße | 1 | 2 | 3 | 4 | 5 | 6 |
| | Tiefenstaffelung | 1 | 2 | 3 | 4 | 5 | 6 |
| | Klangfarbe des Nachhalls | 1 | 2 | 3 | 4 | 5 | 6 |
| 2. Stereo Klangeindruck Die Klänge im Stereobild lassen sich eindeutig zuordnen | Ausgewogenheit der Richtungen | 1 | 2 | 3 | 4 | 5 | 6 |
| | Stabilität | 1 | 2 | 3 | 4 | 5 | 6 |
| | Breite des Hörbildes | 1 | 2 | 3 | 4 | 5 | 6 |
| | Lokalisationsschärfe | 1 | 2 | 3 | 4 | 5 | 6 |
| 3. Transparenz: Alle Details können deutlich differenziert werden | Zuordnung Klangerzeuger – Klang | 1 | 2 | 3 | 4 | 5 | 6 |
| 4. Sound Balance Die einzelnen Geräusche sind gut zum Gesamtsound ausbalanciert | Lautstärkeverhältnisse | 1 | 2 | 3 | 4 | 5 | 6 |
| | Dynamikumfang | 1 | 2 | 3 | 4 | 5 | 6 |
| 7. Gesamteindruck :Eine Zusammenfassung der vorherigen sechs Parameter die den Gesamteindruck der Hörprobe umschreibt. | | 1 | 2 | 3 | 4 | 5 | 6 |

8. Bewertungsskala für subjektive Einschätzungen nach EBU- Tech 3286 (1997)

| | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|

Die Bewertungsskala zur subjektiven Einschätzung stellt für jeden zu bewertenden Parameter sechs Bewertungsstufen zur Verfügung*.

Es handelt sich bei diesen Bewertungsstufen um Ordnungszahlen, bzw. Qualitätsabstufungen.

Daher sollen Urteile nicht als Teilmengen von verschiedenen Stufen oder Zwischenstufen getroffen werden.

Eine Zuordnung der Bewertungsstufen zu subjektiven Empfinden sollte sich an der folgenden Zuordnung Orientieren:

| | Qualität | Eindruck |
|---|--|---|
| 6 | Schlecht Substantielle Technische Mängel / Unbrauchbar | Sehr unzumutbare Fehler |
| 5 | Mangelhaft Nur unter großen Ausnahmen verwendbar oder zu Dokumentarischen Zwecken | Zu viele unzumutbare Fehler |
| 4 | Ausreichend | Eine gewisse Anzahl unzumutbarer Fehler |
| 3 | Gut | Sehr wenige unzumutbare Fehler |
| 2 | Sehr gut | Wahrnehmbare aber zumutbare Fehler |
| 1 | Exzellente | Keine wahrnehmbaren Fehler |

* Die Bewertung wurde dazu dem deutschen Schulnotensystem angepasst, um Verwechslungen zu vermeiden.

Bewertungskriterien

| Hauptmerkmale | Zugeordnete Merkmale | Beispiele zur Beschreibung |
|--|---|---|
| 1. Räumlicher Eindruck Das Klageereignis findet in einem eindeutig definiertem räumlichen Umfeld statt | Homogenität des Raumes Nachhall Akustische Ausgewogenheit Ersichtliche Raumgröße Tiefenstaffelung Klangfarbe des Nachhalls | Hallig / trocken Direkt / indirekt Großer /kleiner Raum |
| 2. Stereo Klangeindruck Die Klänge im Stereobild lassen sich eindeutig zuordnen | Ausgewogenheit der Richtungen Stabilität Breite des Hörbildes Lokalisationsschärfe | Breit / eng Präzise / Unpräzise |
| 3. Transparenz Alle Details können deutlich differenziert werden | Zuordnung von Klangerzeuger zu Klang | Klar / Matschig |
| 4. Sound Balance Die einzelnen Geräusche sind gut zum Gesamtsound ausbalanciert | Lautstärkeverhältnisse Dynamikumfang | Klang zu laut / leise Zu bedämpft / natürlicher Klang |
| 5. Gesamteindruck | Eine Zusammenfassung der vorherigen Parameter die den Gesamteindruck der Hörprobe umschreibt. | |

Definition der Haupt- und Untermerkmale

| | |
|---------------------------|--|
| Homogenität des Raumes | Der subjektive Eindruck das der Raum ein homogenes Ganzes ergibt |
| Nachhall | Der subjektive Eindruck über die Dauer des Nachhalls |
| Akustische Ausgewogenheit | Der subjektive Eindruck des Verhältnisses von direkten und indirekten Schall |
| Ersichtliche Raumgröße | Der subjektive Eindruck von der vorhandenen Raumgröße – real oder künstlich |

| | |
|-----------------------------|--|
| Tiefenstaffelung | Der subjektive Eindruck, dass das Klangbild eine von vorn nach hinten durchsichtige Tiefe besitzt |
| Klangfarbe | Das spezifische Gemisch aus Grundton, Obertönen, Rauschanteilen etc., sowie den zeitlichen Verlauf des Spektrums und der Lautstärke. |
| Klangfarbe des Nachhalls | siehe oben, bezogen auf den Hallanteil |
| Ausgewogenheit der Richtung | Der subjektive Eindruck, dass die Klangquellen im Stereobild so platziert sind, dass ein Gleichgewicht entsteht |
| Stabilität | Der subjektive Eindruck, dass alle Klangquellen an ihren ihnen zugeordneten Positionen verbleiben |
| Lokalisationsschärfe | Der subjektive Eindruck, dass alle Elemente der Darbietung gut im Stereobild positioniert sind |
| Lautstärkeverhältnisse | Der subjektive Eindruck, dass alle Elemente der Darbietung mit angemessener Lautstärke vorhanden sind |
| Dynamikumfang | Der subjektive Eindruck, des Umfangs der dargebotenen Lautstärken |
| Gesamteindruck | Ein subjektiv gewichteter Wert für die gesamte Darbietung, die sich aus der Berücksichtigung der genannten Parameter zusammensetzt |

9. Hörspielskript

3xUnschuld

Kurzhörspiel

von Simon Jaspersen

PERSONAL:

| | | | | |
|----------------------------|---|---------|---|-----------------|
| Marko (18) | - | M | - | Felix Maue |
| Jugendlicher (18) | - | J | - | Deniz Jaspersen |
| Bürgerin (40) mit Kind (4) | - | B und K | - | Helene Matern |
| Wachmann (63) | - | W | - | Oliver Törner |

Normaler Text ist innerer Monolog der Figuren.

Fetter Text ist wörtliche Rede in der Szene.

Kursiver Text sind Regieanweisungen und Technik.

INTRO

Jugendlicher: „Es war nicht meine Schuld.“

Bürgerin: „Was jede Mutter getan hätte.“

Wachmann: „Es waren ja nur Sekunden. Ich...“

Musik: The Notwist – „This Room“

Ansage: **3X Unschuld**

I:

Stimme Jugendlicher: Es war nicht meine Schuld. Nicht meine Schuld. Sicher nicht, ich mein... Wir waren zusammen saufen. Marko und ich. Wie immer. Es war schon etwas heftig zugegangen. Aber dann. Ich kann mich noch erinnern. Deutlich.

(Atmo I)

Wir stehen auf dem Bahnsteig. Es wird schon hell. Wir sind besoffen. Jemand wühlt im Müll. Am Fenster sind Lichter. Und das Rauschen der Stadt.

(Geräusch Flasche)

Plötzlich sieht Marko mich an. Kippe in der Hand sagt er:

Marko: **Ich muss dir... was sagen. Ich kann nicht mehr ich... – es geht schon zu lange. Du musst es wissen.**

Ich und Carla. Wir schlafen mit einander. Schon seit einem Jahr. Wir wollten es dir sagen, aber... Sorry.

Stimme Jugendlicher: Mein Kopf kippt...
Hinter uns kommt die Bahn.

(Bahn LS -> RS)

Mir wird heiß. Es kommt. Von unten. Wie ein böser, roter Ball.
Zerfrisst alles. Er weicht zurück.

Jugendlicher: **(zischend) Du, du... Penner.**

Stimme Jugendlicher: Ich schlage ihm die Bierflasche an den Hals. Er läuft los.
(Klirren)
(Blende)

II.

Stimme Bürgerin: Wir waren auf dem Weg zur Frühkrippe. Martin konnte ausschlafen und ich war übermüdet vom Seminar bei der GAL. Ihre Stricktasche hatte sie am Arm. Ich bin mir doch ganz sicher...
Ihre runde Gestalt. Meine Kleine.
(Atmo II)
Meine Tochter und ich stehen in der Mitte der Plattform.
Die Bahn nach Pinneberg fährt ein.
(Bahn L -> LS)

Kind: **Guck mal Mami. Die Bahn kommt... (R)**

Bürgerin: **Ja, mein Zimtstern.**

Stimme Bürgerin: Plötzlich ist etwas hinter uns.
(leises Klirren LS, Schritte LS -> R)
Dieser Mann läuft vorbei. Knallt gegen Jenny.

Kind: **Auaa! (beginnt zu weinen, R)**

Stimme Bürgerin: Er hat dunkle Haare. Rucksack. Rennt vorbei. Was hat er da vor der Brust? Wo ist ihre Tasche?

Bürgerin: **Komm her Jenny! Hilfe!**

Stimme Bürgerin: Ich ducke mich über meine Tochter. Was jede normale Frau getan hätte.

(Blende)

III:

Stimme Wachmann: Ich will das nicht erinnern. Alle Details. Noch mal und noch mal. Es waren ja nur Sekunden... Was hätte ich...

Man vergisst es nie!

Alleine zu Hause hab ich geweint. Die Bilder im Fernsehen. Schlimmer als London. Er war so jung. Noch nicht bereit. Ich kann es nicht abstellen...

(Atmo III)

4:48 Uhr. Ich mache Durchsage.

Wachmann: **Achtung auf Gleis zwei ist eingef...**

Bürgerin: **Komm her Jenny! Hilfe! (RS)**

Wachmann: Da kommt er gelaufen.

(Schritte R -> L)

Wachmann: **Hey, was machen Sie denn da?**

Bleiben sie stehen.

Stimme Wachmann: Er hört ja nicht. Hält seine Arme. Vor der Brust. Hinter ihm schreit die Frau. Panik bekomme ich. Er sieht mich – und nimmt reiß aus. Ich renne auch.

(Schritte M)

Wachmann: **Sofort stehen bleiben...**

Frau: **Halten Sie ihn auf! Er ist gefährlich.** (R -> RS)

Stimme Wachmann: Der Rucksack. Da guckt etwas raus. Was versteckt er da? Er ist so gebückt. Läuft auf die Menschen zu. Gefährlich! Attentäter! Ich ziehe das Elektroimpulsgerät. Die Bahn. Wenn er sie erreicht... Was soll ich tun?
(Bahn L -> LS)

Wachmann: **Stehen bleiben oder ich schieße...!**

(Taserschuss) (Schreien L, Bahn)
Stimme Wachmann: Habe ich ihn getroffen? Weiß nicht. Er taumelt. Und verpasst den Bahnsteig.
Ich habe rechtens gehandelt.
(Blende)

OUTRO

Das Hamburger Polizeigesetz ist in wesentlichen Punkten geändert worden. Zu den stark erweiterten Rechten der Polizei gehört ab nun auch der Einsatz von „Distanz – Elektroimpulsgeräten“. Dabei handelt es sich um Schockwaffen, die eine Spannung von 50.000 Volt in die Körper der Opfer übertragen.

Amnesty International zu Folge sind seit dem Jahr 2001 in den USA über 70 Menschen an den Folgen dieser Einsätze gestorben.

Musik: The Notwist – „Neon Golden“

Absage

10. Lageplan U-Bahnhaltestelle Hamburg-Jungferstieg

