Optimierter Stereo-Downmix von 5.1-Mehrkanalproduktionen (An optimized Stereo-Downmix of a 5.1 multichannel audio production)

Bernfried Runow*, (Jörg Deigmöller**)

- * runow@b-public.de
- ** deigmoeller@irt.de

Kurzfassung

Für die Verwendung des immer beliebteren Mehrkanaltons bei audiovisuellen Inhalten muss in den meisten Fällen die Kompatibilität zu bisher verwendeten Tonformaten wie Zweikanal-Stereo oder Mono gewährleistet sein. Mit einem automatischen Stereodownmix kann diese Kompatibilität effizient erreicht werden. Jedoch bringt ein automatisches Matrizierungsverfahren Probleme mit sich: Klangfarbe, Räumlichkeit, Lautstärkebalance und Abbildung können verändert und damit die technische Qualität sowie künstlerische Aspekte der Produktion negativ beeinflusst werden.

Am IRT wurde ein neues Downmixverfahren entwickelt, das auf umfangreichen Forschungsarbeiten zur Qualität solcher Verfahren beruht. Mit dem entstandenen Algorithmus können die beim Downmix auftretenden, störenden Effekte verhindert oder wirksam kompensiert werden.

1. Einleitung

Ein Downmix ist ein manuelles oder automatisches Verfahren, bei dem die akustische Information eines mehrkanaligen Tonformats auf eine geringere Anzahl diskreter Kanäle zusammengefasst wird. In der Praxis wird ein Downmix meist angewandt, um von einem Mehrkanalton eine Zweikanal-Stereo-Fassung (Mono-Fassung) zu erstellen und so die Abwärtskompatibilität des Mehrkanalformats zu sichern.



Abb. 1: Prinzip eines Downmixverfahrens

Der Anspruch an einen Stereodownmix ist eine Zweikanal-Stereofassung, welche ein sinnvolles Klangabbild des Mehrkanaltons wiedergibt. Dabei ist die akustische Information in dem Bereich zwischen den beiden verbleibenden Lautsprechern so unterzubringen, dass sie beim Hörer möglichst gleich oder zumindest ähnlich wahrgenommen wird.

Das unumgängliche Matrizieren der einzelnen Kanäle des Mehrkanaltons beim Downmix bringt einige Probleme mit sich. Laufzeitunterschiede kohärenter Signalanteile in den einzelnen Kanälen führen zu Kammfiltereffekten, unterschiedliche Additionseigenschaften akustischer und elektrischer Signale sowie kohärenter und inkohärenter Signalanteile haben

eine Änderung der Lautstärkebalance und der Position von Phantomschallquellen bei der Downmixversion zur Folge. Solch ein unbeabsichtigter und nicht kontrollierbarer Eingriff in das Klangbild kann die technische Qualität sowie die ursprüngliche Intention des Tongestaltenden negativ beeinflussen.

2. Bisher zur Verfügung stehende Downmixverfahren

Es wird zwischen passiven und aktiven Downmixverfahren unterschieden. Passive Downmixverfahren fassen den Mehrkanalton ohne eine vorangehende Signalanalyse, also mit festen Einstellungen und statischen Parametern zu zwei Kanälen zusammen. Sie arbeiten unabhängig von der Beschaffenheit des Ausgangsmaterials. Aktive Downmixverfahren zeichnen sich durch eine dynamische Anpassung der Parameter und eine intelligente Bearbeitung aus, um ein bestmögliches Ergebnis zu erzielen.

2.1. Passive Downmixverfahren

Die beiden bekanntesten passiven Downmixverfahren sind die Downmix-Empfehlung der ITU und ein Verfahren von Dolby, das Teil des Dolby Pro Logic II Codecs ist.

2.1.1. ITU Downmix

In der Veröffentlichung "BS.775-2" der ITU wird neben einer Empfehlung zur Lautsprecherpositionierung und der optimalen Abhörposition (engl.: Sweet Spot) auch ein passives Downmixverfahren vorgeschlagen, das die Abwärtskompatibilität sicherstellen soll. [1]

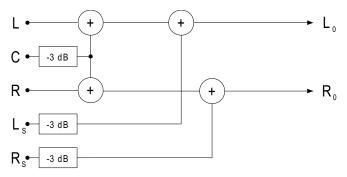


Abb. 2: Blockschaltbild des Downmixes nach ITU BS.775-2

2.1.2. Dolby Pro Logic II

Dolby bezweckt mit seinem Codec nicht nur einen Downmix auf die zwei Kanäle L_t (left total) und R_t (right total), sondern ermöglicht auch die Wiederherstellung eines Mehrkanaltons aus L_t und R_t . Die Information zur Trennung der Kanäle beim Upmix entsteht beim Downmix durch das Zumischen der Surroundkanäle mit einer Phasendrehung.

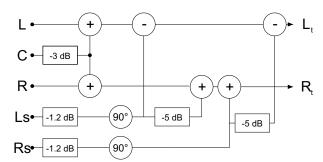


Abb. 3: Blockschaltbild des Dolby Pro Logic II Encoders

2.2. Aktive Downmixverfahren

Aktive Downmixverfahren zeichnen sich durch eine dynamische Anpassung der Parameter abhängig vom Ausgangsmaterial aus.

2.2.1. Lexicon Logic7

Das von David Griesinger entwickelte aktive Verfahren beinhaltet neben einem Downmix auch die Möglichkeit zum Upmix. Griesinger legt bei seinem Downmixverfahren besonders Wert auf die Erhaltung der effektiven Energie, eine möglichst geringe Verfälschung der Phantomschallquellen sowie die dynamische Anpassung der Dämpfungsparameter des Centerkanals und der Surroundkanäle. [2]

2.2.2. Coding Technologies

Ausgehend von der Downmixempfehlung der ITU wird bei dem noch in der Entwicklung befindlichen Verfahren der Energiegehalt nach der Addition in 28 Frequenzbändern analysiert. Anschließend werden zu starke Anhebungen und Absenkungen korrigiert. Die Bildung von Kammfiltereffekten kann auf diese Weise eingeschränkt werden.

Leider hat die Firma Coding Technologies nach der Übernahme von Dolby das Projekt vorläufig eingestellt.

2.3. Untersuchungen von Downmixverfahren

Am Institut für Rundfunktechnik wurden in den vergangenen Jahren mehrfach Tests und Hörversuche mit den erwähnten Downmixverfahren durchgeführt. Neben den Vergleichen der einzelnen Downmixversionen untereinander wurden in den Hörversuchen auch die subjektiv wahrnehmbaren Unterschiede zu einer von Hand erstellten Stereoversion untersucht.

Die Ergebnisse dieser Untersuchungen offenbaren, dass aktive Downmixverfahren den passiven überlegen sind. Verfahren, die einen späteren Upmix durch eine Phasenveränderung beim Downmix ermöglichen, müssen Einschränkungen bei der erstellten Zweikanalversion hinnehmen. Auch zeigen die Auswertungen, dass die von Hand erstellte Stereo-Mischung im Durchschnitt nicht besser bewertet wurde als die automatisch erstellten Downmixversionen. Genreabhängig wurde die Hand-Mischung sogar schlechter bewertet. [3][4] Dies darf jedoch nicht darüber hinwegtäuschen, dass sich der Handmix im Gegensatz zu einem automatischen Downmix durch künstlerische Individualität auszeichnet und daher, wenn er zur Verfügung steht, bevorzugt werden sollte. Dennoch sind aktive Downmixverfahren hinsichtlich ihrer technischen Qualität in der Lage, mit einer von Hand erstellten Mischung zu konkurrieren und stellen deshalb eine sinnvolle Alternative dar, falls die Mischung der Zweikanalversion von Hand nicht möglich ist.

Bei einer Umfrage nach dem Hörversuch wurden die Probanden befragt, welche Eigenschaften des Klangbildes bei einem Downmix erhalten bleiben sollten. An erster Stelle stand die Klangfarbe, gefolgt von der Lautstärkebalance und der Räumlichkeit.

3. Optimierung des Stereo-Downmixes

Mit den gewonnenen Erkenntnissen der genannten Evaluierung wurden am Institut für Rundfunktechnik Lösungen für die beim Downmix auftretenden Probleme entwickelt.

¹ Die 28 Frequenzbänder sind von der Bark-Skala (vgl. Zwicker, Feldtkeller 1967) abgeleitet, die dem menschlichen Hörverhalten angepasst ist.

Ein Großteil der Änderungen in der Klangfarbe, die als besonders störend empfunden werden, entsteht durch Kammfiltereffekte. Ein wesentlicher Ansatzpunkt ist daher die Kompensation solcher Effekte. Ausgehend von der ITU-Empfehlung wird an allen Summationspunkten eine Kammfilterkompensation integriert. Der Ausgleich zu starker Anhebungen kohärenter Signalanteile sowie eine größere Abbildungsstabilität von Phantomschallquellen werden durch eine Energiekorrektur erreicht. Außerdem sollen die Parameter zur Dämpfung der Surroundkanäle und des Centerkanals variabel bleiben, so dass diese beispielsweise abhängig vom Genre sinnvoll eingestellt werden können. Abb. 4 zeigt den als Blockschaltbild visualisierten Ansatz des IRT-Downmixes.

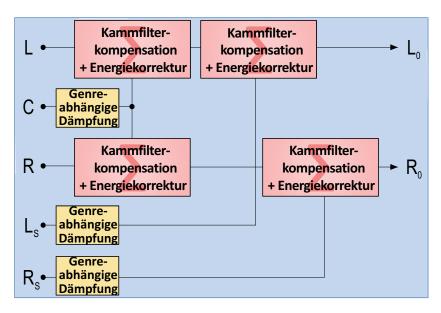


Abb. 4: Ansatz des IRT-Downmixes als Blockschaltbild

3.1. Klangfarbe

Die Änderungen in der Klangfarbe werden unterbunden, indem bei jedem Additionsvorgang der Betrag der resultierende Summe IST in sehr schmalbandigen Frequenzbändern auf Kammfilter untersucht wird. Um eine Anhebung oder Absenkung detektieren zu können, wird ein Richtwert benötigt, mit dem der Wert eines jeden Frequenzbandes verglichen werden kann. Der Richtwert sollte den Betrag aufweisen, den die Summe IST hätte, läge keinerlei Beeinflussung durch einen Kammfilter vor. Diesen Betrag erhält man, wenn man die beiden zu addierenden Signale erst dekorreliert - also in einen rechten Winkel zueinander bringt - und dann addiert. Das Ergebnis, der SOLL-Wert, entspricht der energetischen Summe, die unabhängig von der Phasenlage der beiden zu addierenden Signale und maximal 3 dB größer als der Betrag der zu addierenden Signale ist.

$$IST = |A + B|, (1)$$

$$SOLL = \sqrt{A^2 + B^2}.$$
 (2)

Wird ein Kammfilter detektiert - liegt also der IST-Wert über oder unter dem SOLL-Wert - so kann eine Bearbeitung mit dem Ziel der Kompensation des Kammfiltereffekts eingeleitet werden. Optisch gesehen ist das Ziel, der für Kammfilter charakteristischen Wellenform im Spektrum entgegen zu wirken. Hat die Summe durch den Kammfilter eine Absenkung erfahren, so wird sie verstärkt, hat sie eine Anhebung erfahren, wird sie gedämpft (Abb. 5).

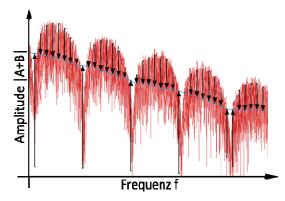


Abb. 5: Grafische Veranschaulichung der Kammfilterkompensation

Um die Stärke der Kompensation in Bezug auf die Veränderung durch einen Kammfilter bewerten zu können, eignet sich die grafische Darstellung in einem Übertragungsdiagramm. Am Eingang wird der Pegel des IST-Werts in Bezug auf den SOLL-Wert, am Ausgang die Summe nach der Kompensation in Bezug auf den SOLL-Wert angezeigt. Es entsteht eine Kennlinie, welche die Stärke der Kompensation in Bezug auf die Stärke der Anhebung bzw. Absenkung angibt.

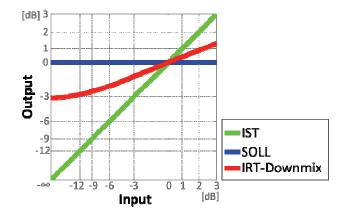


Abb. 6: Übertragungsdiagramm mit Kammfilterkompensations-Kennlinien (Pegel [dB] in Bezug auf den SOLL-Wert)

Wird das Signal nicht bearbeitet, so erhält man die IST-Kennlinie. Die vollständige Kompensation des Kammfilters stellt die SOLL-Kennlinie dar. Eine vollständige Kompensation führt in der Praxis jedoch zu einer Vielzahl von Artefakten, so dass der IRT-Downmix einen Kammfilter nur annähernd ausgleicht (Kennlinie IRT-Downmix). Der Schnittpunkt der drei Kennlinien repräsentiert den Fall, bei dem IST- und SOLL-Wert gleich sind. Tritt dieser Fall ein, liegt keine Veränderung durch einen Kammfilter vor, so dass keine Bearbeitung notwendig ist.

Die eigentliche Kompensation wird durch die Multiplikation des IST-Werts mit einem reellen Faktor x erreicht (siehe Abb. 7, Methode 1). Löschen sich die beiden Signale A und B allerdings vollständig aus, so ist es nicht möglich, den resultierenden IST-Wert an den SOLL-Wert anzugleichen. Enthält der IST-Wert keine Information mehr, so bleibt nur die Möglichkeit, auf die beiden zu addierenden Signale zurückzugreifen. Diese müssen bei einer kompletten Auslöschung kohärent und um genau 180° phasenverschoben sein sowie den gleichen Betrag aufweisen. Verändert man den Betrag des ersten zu addierenden Signals mit einem rationalen Faktor x, so kann die vollständige Auslöschung verhindert werden (siehe Abb. 7, Methode 2).

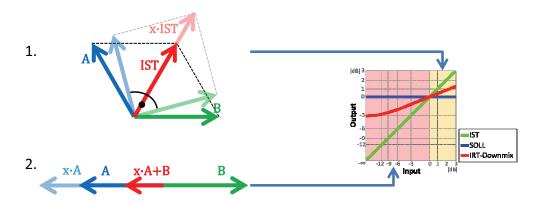


Abb. 7: Kammfilterkompensation des IRT-Downmixes im Detail

Auslöschungen können mit Methode 2, Anhebungen mit Methode 1 korrigiert werden. Die Bearbeitung von Absenkungen ist mit beiden Methoden möglich. Es muss also ein Übergang von Bearbeitungsmethode 1 zu Methode 2 geschaffen werden. Damit es zu keinem Phasensprung kommt, ist es wichtig, dass beide Methoden am Punkt des Übergangs die gleiche Phase liefern. Dies ist nur an einer einzigen Stelle im Übertragungsdiagramm der Fall, nämlich dann, wenn keine Korrektur notwendig ist, d.h. wenn der IST-Wert am Eingang gleich dem SOLL-Wert ist (siehe Abb. 7, Übertragungsdiagramm).

3.2. Lautstärkebalance

Bei der Wiedergabe des 5.1-Mehrkanaltons addieren sich die einzelnen Kanäle akustisch. Bei einem Downmix werden sie hingegen elektrisch addiert. Auf Grund von unterschiedlichen Additionseigenschaften von kohärenten und inkohärenten Signalen bei der akustischen und elektrischen Addition kommt es beim Downmix deshalb zu einer Veränderung der Lautstärkebalance zwischen kohärenten und inkohärenten Signalanteilen.

Die elektrische Addition kohärenter Signale liefert einen um bis zu 3 dB höheren Schalldruckpegel als die Wiedergabe des kohärenten Signals über zwei getrennte Lautsprecher (siehe Abb. 8). Im Gegensatz dazu wird bei der akustischen und elektrischen Addition inkohärenter Signale keine Veränderung des resultierenden Schalldruckpegels wahrgenommen.

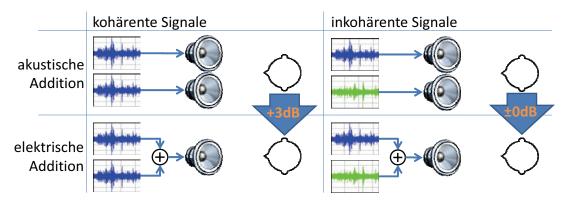


Abb. 8: Akustische und elektrische Signaladdition

Die Folge ist, dass kohärente und phasengleiche Signalanteile - zum Beispiel Phantomschallquellen, die mit Hilfe von Pegeldifferenzen gebildet wurden - nach dem Downmix im Vergleich zu inkohärenten Signalanteilen einen höheren Pegel aufweisen.

Bei einer aus Pegeldifferenzen gebildeten Phantomschallquelle ist die Änderung ihres Pegels beim Downmix zudem abhängig von ihrer Auslenkung. Der Pegel einer Phantomschallquelle, die genau zwischen L/C bzw. C/R angeordnet ist, wird dabei am stärksten verändert.

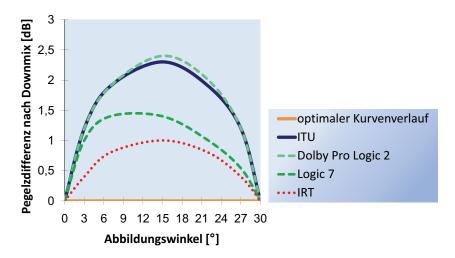


Abb. 9: Pegeländerung einer Phantomschallquelle nach dem Downmix

Wird die energetische Summe als Richtwert bei der Kammfilterkompensation verwendet, so werden auch alle zu starken Anhebungen kohärenter Signalanteile korrigiert. Der im Vergleich zur akustischen Addition um bis zu 3 dB höhere Pegel wird gedämpft, so dass die Pegeländerung deutlich geringer ausfällt (siehe Abb. 9, IRT-Downmix).

3.3. Abbildung

Nicht nur der Pegel einer Phantomschallquelle ändert sich beim Downmix, sondern auch deren Auslenkung aus der Mitte. Eine durch Pegeldifferenzen gebildete Phantomschallquelle im Frontbereich wird durch den Downmix in der L/R-Stereobasis weiter nach außen hin verschoben. Auch in diesem Fall wirkt sich die Angleichung an die energetische Summe positiv aus. Der zu hohe Pegel im linken bzw. rechten Downmixkanal wird gedämpft, so dass die Winkeländerungen auf unter 2° korrigiert werden können (siehe Abb. 10).

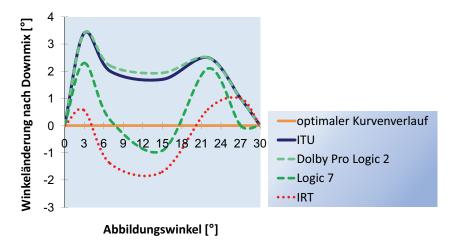


Abb. 10: Winkeländerung einer Phantomschallquelle nach dem Downmix

Phantomschallquellen, die mit Hilfe von Laufzeitdifferenzen gebildet wurden, stellen eine noch größere Herausforderung für ein Downmixverfahren dar. Die Summe kohärenter,

jedoch zeitlich verschobener Signale erfährt bei der Addition keine über den gesamten Frequenzbereich konstante Pegelanhebung. Ein Kammfilter entsteht, der die Klangfarbe verändert. Auf Grund der Pegelabsenkung des Centerkanals um 3 dB kommen nach dem Downmix zusätzliche Pegeldifferenzen zwischen L und R hinzu, die gegenläufig zu den Laufzeitdifferenzen wirken können. Die Abbildungseigenschaften der Phantomschallquelle werden dadurch deutlich verschlechtert.

Durch die Verwendung der Kammfilterkompensation, die sich an der energetischen Summe orientiert, kann das Downmixergebnis aber auch in diesem Fall erheblich verbessert werden.

4. Evaluierung des optimierten Downmixverfahrens

Zur Evaluation des optimierten IRT-Downmixes wurde dieser in einem Hörversuch mit dem Downmix nach der ITU-Empfehlung und einer von Hand erstellten Stereomischung verglichen. An dem Versuch nahmen 12 Mitarbeiter des Instituts für Rundfunktechnik und des Bayerischen Rundfunks teil. Zur Durchführung wurde ein Paarvergleich (A-B-Vergleich) gewählt, da weder der Mehrkanalton noch eine der Downmixversionen als Referenz angesehen werden können. Dem Probanden standen damit zwei Stereoversionen A und B gegenüber, die er hinsichtlich ihrer subjektiv wahrgenommenen Qualität vergleichen sollte. Um ein möglichst breit gefächertes Spektrum abzudecken, wurden Hörbeispiele aus den Genres Klassik, Chor, Jazz, Pop, Radiokunstproduktion und TV-Show gewählt sowie drei synthetische Hörbeispiele, die gezielt Probleme in den Bereichen Klangfarbe, Lautstärkebalance und Abbildung hervorrufen.

Die grafische Auswertung des Hörversuchs wurde mit dem Medianwert und den Quartilen realisiert. Abb. 11 und Abb. 12 zeigen die Ergebnisse des Hörversuches. Der Wert "0" im Diagramm sagt aus, dass die beiden Versionen als gleichwertig beurteilt wurden. Bei Werten im positiven, oberen Bereich des Diagramms wurde der IRT-Downmix als besser befunden, bei Werten im negativen, unteren Bereich wurde der Downmix, der anhand der ITU-Empfehlung entstand, besser beurteilt.

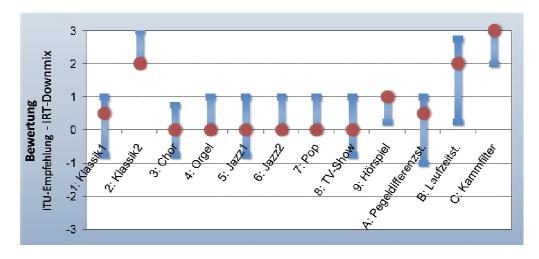


Abb. 11: Grafische Auswertung des Paarvergleichs: IRT-Downmix - ITU-Empfehlung

Die Grafik zeigt deutlich die positive Auswirkung der Kammfilterkompensation des IRT-Downmixes. Gerade bei den Klassik-Beispielen, dem Hörspiel und den synthetischen Hörbeispielen (A, B, C) ist offensichtlich eine Qualitätssteigerung wahrzunehmen.

Ein Vergleich mit einer von Hand erstellten Stereoversion ist nur bedingt möglich, da der künstlerische Wert bei einem Hörversuch oft nur beschränkt erfasst wird, und kann daher lediglich hinsichtlich der technischen Qualität ein sinnvolles Ergebnis liefern. Dennoch ist es für die Einschätzung eines automatischen Verfahrens wichtig, ob der Hörer einen Unterschied wahrnehmen kann.

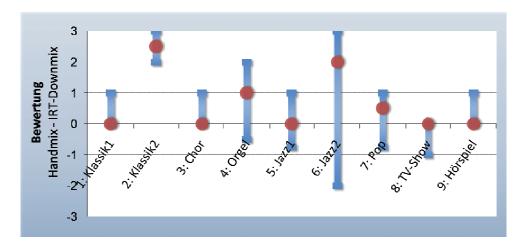


Abb. 12: Grafische Auswertung des Paarvergleichs: IRT-Downmix - Handmischung

Die Ergebnisse zeigen, dass ein optimiertes Downmixverfahren, wie der am IRT entwickelte Algorithmus, qualitativ hochwertige Resultate liefert, die keine wahrnehmbaren Schwächen gegenüber einer von Hand erstellten Stereoversion aufweisen. Gegenüber passiven Verfahren wie der ITU-Empfehlung liegt der IRT-Downmix gerade bei kritischem Ausgangsmaterial klar im Vorteil.

Das vorgestellte Verfahren ermöglicht einen Downmix, bei dem die bisher störenden Veränderungen in der Klangfarbe, Lautstärkebalance und Abbildung auf ein Minimum reduziert werden. Auch bei der Räumlichkeit und Transparenz wirkt sich die Optimierung wahrnehmbar positiv aus. Damit steht in Zeiten, in denen immer mehr Endformate bedient werden müssen, ein zuverlässiges Werkzeug für die Herstellung eines Downmixes zur Verfügung, das dann eine sinnvolle Anwendung findet, wenn die Mischung der Zweikanalversion von Hand nicht möglich ist. Mit einer Echtzeit-Implementierung kann der Tongestaltende schon bei der Produktion des Mehrkanaltons den Downmix gegenhören und auf diese Weise das Ergebnis des Downmixes kontrollieren.

5. Quellenverzeichnis

- [1] ITU-R BS.775-2: "Multichannel Stereophonic Sound System With and Without Accompanying Picture" (*Geneva*, 1992-1994)
- [2] Griesinger, David: "Surround From Stereo", http://www.davidgriesinger.com/surround from stereo2.ppt, 2003
- [3] Deigmöller, Jörg: "Subjektive und objektive Untersuchungen zum 2.0-Downmix von 5.1-Mehrkanal-Produktionen (Diplomarbeit)" *IRT München, 2006*
- [4] Keil, Martin: "Realisierung eines Stereo-Downmixes aus einem 5.1 Signal für Anwendungen im Rundfunk (Diplomarbeit)" *IRT München*, 2007