

20. TONMEISTERTAGUNG 1998

Ein virtueller Surround Sound Abhörraum im Ü-Wagen

*Ulrich Horbach, Renato Pellegrini (Studer, CH-8105 Regensdorf),
Uwe Felderhof, Günther Theile (Institut für Rundfunktechnik, 80939 München)*

Es wird ein virtueller Abhörraum für Mehrkanalton vorgestellt. Das Verfahren basiert auf einem speziellen Faltungsalgorithmus für die Auralisation, und es berücksichtigt die Bedeutung der spontanen Kopfdrehungen durch Einbeziehung eines Head-Tracking Systems. Es ist in der Lage, die binauralen Signale des drehbaren Kunstkopfes (Neumann KU 100) vollständig zu reproduzieren: Sämtliche relevanten Außenohrübertragungsfunktionen werden in einem gewünschten realen Abhörraum am Referenzort einer realen 3/2-Stereo Lautsprecheranordnung gemessen und stehen in einer Datenbank für die Faltung zur Verfügung. Im Gegensatz zu anderen Verfahren wird der virtuelle Abhörraum nicht synthetisch erzeugt, sondern er ist „geklont“, mit der realen Abhörsituation am gemessenen Ort identisch. - Der Tonmeister ist damit in der Lage, einen hochwertigen Abhörraum einschließlich der gewohnten Lautsprecher „mitzunehmen“, beispielsweise in den Ü-Wagen. Er ist natürlich auch in der Lage, die Abhörsituation durch Knopfdruck zu wechseln. Mögliche Anwendungsbereiche und Grenzen des Verfahrens werden diskutiert.

A virtual surround sound listening room in an OB van

A virtual surround sound listening room is introduced. The system is based on the so-called “Binaural Room Scanning” (BRS) method. It is able to reproduce the binaural signal of the dummy head (Neumann KU 100) completely. Each of the relevant HRTFs is measured in a real listening room at the reference point of the 3/2-stereo loudspeaker arrangement and stored in a data bank, applicable for convolution with the source signal. In contrast to the usual method of sound field calculation (e.g. ray-tracing), the BRS method does not generate synthetically the virtual room along with corresponding sound sources, rather, the real situation is “cloned”, a virtual duplicate of the listening room at the measuring point. Equipped with a head tracking system the BRS processor enables listening in a virtual room with almost perfect fidelity of reality. The Tonmeister principally gets the opportunity to bring along (e.g. into the OB van) a high quality listening room as well as his familiar loudspeakers. Furthermore, he is able to switch over and to compare different listening conditions, for example studio with home reproduction, or film sound reproduction in a big cinema with television sound reproduction. Possible areas of application and limits of the system are discussed.

1 Einleitung

Verfahren zur Herstellung einer virtuellen auditiven Umgebung werden sowohl im Consumer-Bereich als auch für professionelle Anwendungen eine immer größere Bedeutung erlangen, so z.B. in der Studioteknik, für Telekonferenz- und Telepräsenzsysteme, für Mensch-Maschine-Kommunikation, raumakustische Synthese und Analyse, aber auch für Multimedia (Computerspiele, „Cyberspace“). Aufgabe dieser sogenannten Auralisationsverfahren ist es, für den Hörer eine künstliche auditive Scheinwelt zu schaffen, beispielsweise die scheinbare Anwesenheit in einem virtuellen, schallreflektierenden Raum.

Im vorliegenden Beitrag wird ein neu entwickeltes Gerät vorgestellt, welches es dem Tonmeister ermöglicht, Surround- Tonaufnahmen „virtuell“ über Kopfhörer abzuhören, derart, dass die Wiedergabe weitgehend der in einem existierenden Tonstudio entspricht, dessen Daten vorab mittels binauraler Messungen eingespeichert wurden. Eine anstehende Mehrkanal-Abmischung kann somit über Kopfhörer z.B. im Ü-Wagen durchgeführt werden, d.h. in einer akustisch ungünstigen Umgebung, wo der Aufbau einer 3/2 Lautsprecher-Surroundanordnung gemäss dem aktuellen Standard zu aufwendig wäre. Die erreichte hohe Qualität der klangfarbentreuen Wiedergabe von Original-studio und -lautsprechern und der spontanen Vornelokalisation ist vor allem zwei Faktoren zu verdanken. Einerseits ist das System in der Lage, spontane Kopfdrehbewegungen des Hörers einzubeziehen, welche mittels eines „Headtrackers“ detektiert werden. Des weiteren konnte durch Einsatz von digitalen Signalprozessoren und Speicherbausteinen neuester Generation die dynamische Umschaltung kompletter Raumübertragungsfunktionen, abhängig von der Hörerposition, bei vertretbaren Kosten und Abmessungen implementiert werden, so dass alle relevanten Informationen unverfälscht übertragen werden. Beide Punkte werden im folgenden näher erläutert.

2 Virtuelle Räume: Modell- und Datenbasierte Verfahren

Die Nachbildung einer akustischen Umgebung erfolgt üblicherweise nach dem in Abb. 1 dargestellten Prinzip. Im Modell wird auf eine Datenbank von Aussenohrübertragungsfunktionen $HRTF(L,R)$ zugegriffen, die in einer reflexionsarmen Umgebung vorab ermittelt wurden. Abhängig vom Einfallswinkel einer Schallquelle werden hierbei die frequenzabhängigen Schalldruckverläufe jeweils am linken und rechten Ohr eines Kunstkopfs oder einer Person gemessen und katalogisiert. Mit Hilfe einschlägiger Raumsimulationssoftware können Winkel ($\alpha_1.. \alpha_n$) und Laufzeiten ($t_1..t_n$) von n eintreffenden diskreten frühen Reflexionen ermittelt werden. Je nach Winkel werden den Echos passende HRTF-Paare aus der Datenbank zugewiesen, die zusätzlich noch mit Dämpfungsfaktoren und einfachen Filtern zur Nachbildung der Absorption an Wänden kaskadiert werden können. Die beschriebenen Parameter können sich dynamisch ändern, abhängig von der Position der Quelle, und dem aktuellen Drehwinkel des Kopfes des Hörers, welcher mit einem „Headtracker“ detektiert wird. Die späten Reflexionen werden richtungsunabhängig mit Nachhallprozessoren zugefügt (Abb. 1). Die Bedeutung der Tatsache, dass zur Erzielung einer plausiblen virtuellen Umgebung sowie stabilen Vorneortung die Berücksichtigung der Kopfdrotation (spontaner Kopfdrehbewegungen) unerlässlich ist, wird in zwei getrennten Beiträgen dieser Tagung ausführlich erklärt und soll hier nur noch einmal hervorgehoben werden [1,2].

Offensichtliche Vorteile eines solchen modellbasierten binauralen Raumsimulators sind die Möglichkeit, parametergesteuert beliebige, auch bewegte Quellenpositionen zu realisieren, und die sehr einfache Skalierbarkeit im Hinblick auf die verfügbare Rechenleistung. Sowohl die Anzahl der berechneten Echos, als auch die Komplexität (der Filtergrad) der eingesetzten HRTFs können reduziert werden, ohne dass notwendigerweise drastische Einbussen des räumlichen Höreindrucks auftreten. Dieser Vorteil ist zugleich auch Nachteil: In der Regel werden so stark vereinfachte Modelle verwendet, dass die Realitätsnähe verloren geht. Strebt man eine authentische Abbildung eines Hörereignisses, etwa eines Tonstudios mit dort installierten Lautsprechern an, so nimmt der Aufwand extrem zu. Viele der heute vermarkteten Systeme verzichten daher aus Aufwandsgründen auf die head tracking Einrichtung, die jedoch, wie gesagt, für die räumliche Wahrnehmung eine hervorragende Bedeutung hat.

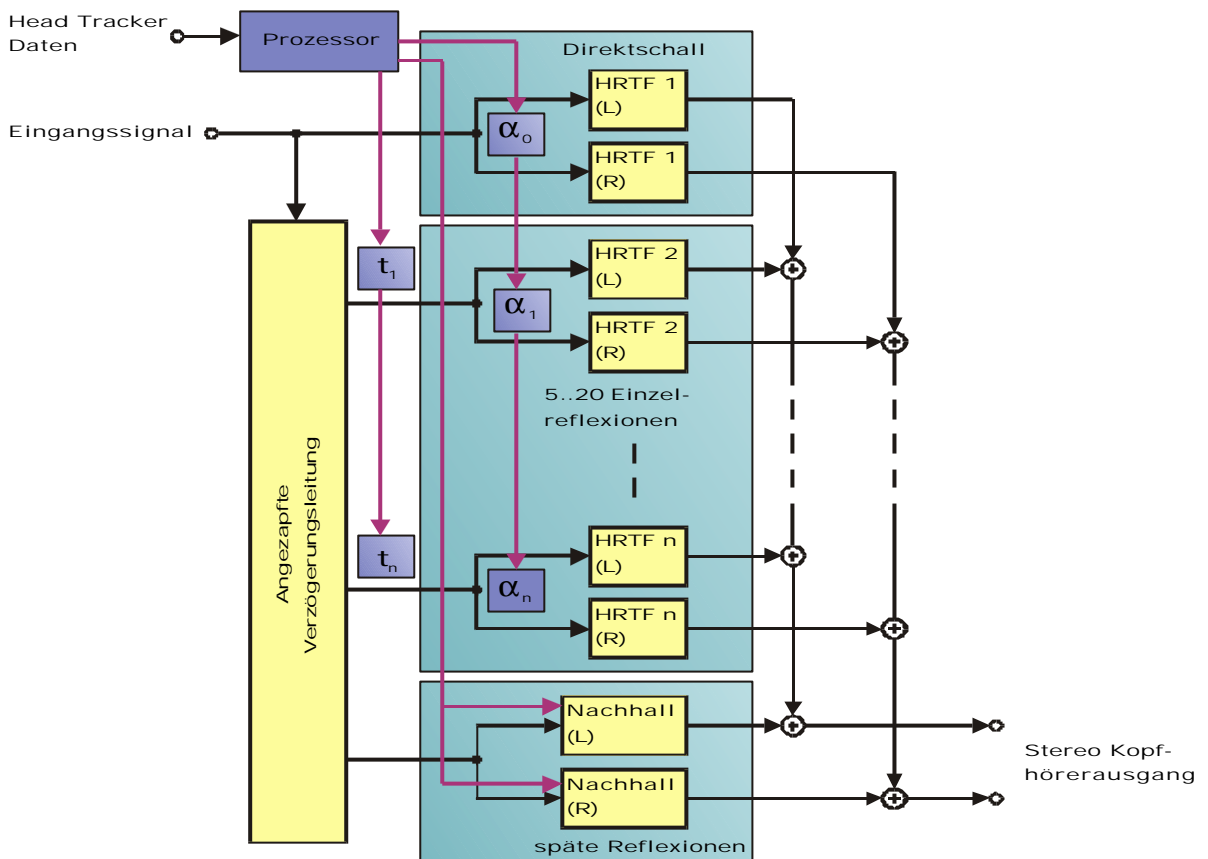


Abb. 1: Modell einer binauralen Raumsynthese

Binaurale Raumimpulsantwort

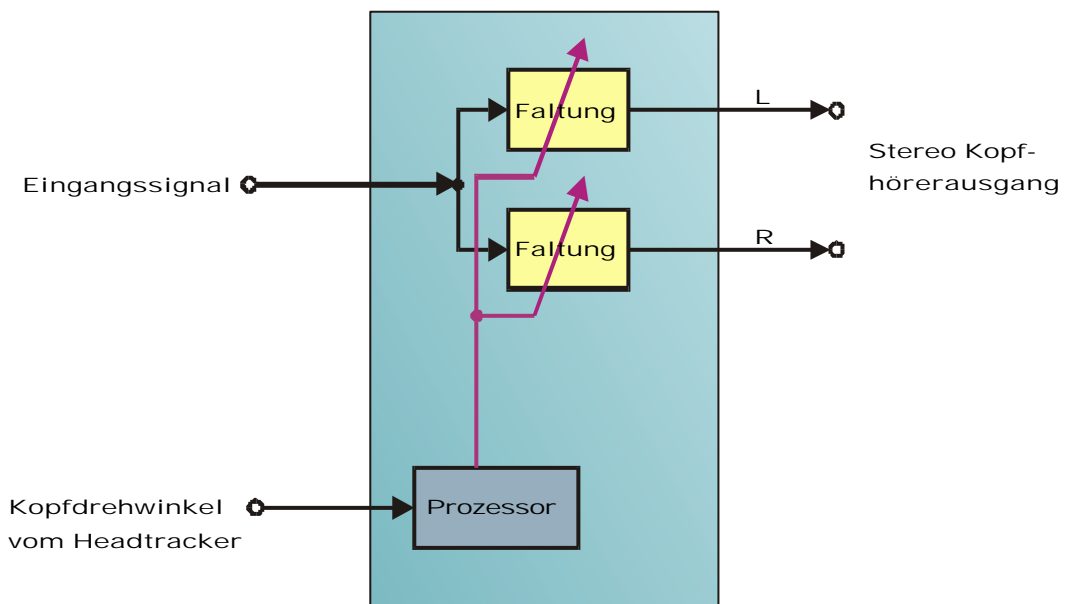


Abb. 2: Datenbasierter Ansatz

Wir können nun davon Gebrauch machen, dass in der hier geforderten Anwendung keine bewegten, sondern nur eine begrenzte Anzahl fester Quellen simuliert und wiedergegeben werden müssen, nämlich die Lautsprecher der Surroundanordnung. Hier empfiehlt sich der sogenannte „datenbasierte“ Ansatz. Wir verzichten auf eine Raumsynthese und verwenden statt dessen direkt die Messdaten. Ein kompletter Satz der binauralen Impulsantworten wird für jeden der Lautsprecher mit Hilfe eines exakt diffusfeldentzerrten Kunstkopfes (z.B. Neumann KU 100) gemessen. Der Datensatz umfaßt alle relevanten Ausrichtungen des Kunstkopfes und steht in einer Datenbank für die dynamische Generierung der binauralen Signale zur Verfügung. Diese werden im Prozessor durch Faltung der binauralen Impulsantworten mit dem Quellsignal erzeugt (Abb. 2). Im Gegensatz zum ursprünglichen Konzept der Schallfeldberechnung erzeugt das „BRS“ (Binaural Room Scanning) genannte Verfahren den virtuellen Raum und die darin angeordneten Schallquellen also nicht synthetisch, sondern sie sind „geklont“, mit der realen Situation am gemessenen Ort identisch. Es berücksichtigt darüber hinaus durch Einbeziehung eines Head-Tracking Systems die hervorragende Bedeutung der spontanen Kopfdrehungen und gewährleistet dadurch eine bisher unbekannte Realitätstreue beim Hören in einem virtuellen Raum.

Auf einige Details der technischen Realisierung soll im nächsten Abschnitt eingegangen werden.

3 Der BRS- Prozessor

Der BRS Prozessor ist in einem 19“ Gerät (eine Höheneinheit) untergebracht. Die gesamte Signalverarbeitung (Speicher, DSPs) befindet sich auf einem einzigen Aufsteckboard im Gerät. Ein vereinfachtes Blockschema zeigt Abbildung 3. Die notwendigen Faltungsoperationen werden, identisch für alle Eingangskanäle, als „schnelle Faltungen“ im Frequenzbereich durchgeführt. Zur Erzielung einer kurzen Latenzzeit (Signallaufzeit) <6msec wurden die Längen der Ein- und Ausgangspuffer und damit der FFTs entsprechend klein gewählt. Um trotzdem lange Raumimpulsantworten verarbeiten zu können, müssen die komplexwertigen Produkte des Signalspektrums innerhalb einer Periode mehrfach mit entsprechend kurzen verzögerten Abschnitten der Raumspektren gerechnet und aufsummiert werden, ähnlich wie bei einem FIR Filter. Diese Operation wurde in Abb. 3 als „komplexe Faltung“ bezeichnet. Die Raumdaten sind in Schritten von (2° .. 6°) gespeichert. Damit bei der

Kopfbewegung keine Störgeräusche oder sonstige Artefakte auftreten, müssen die Spektren gesteuert vom detektierten Kopfdrehwinkel in einem separaten „Prozessor“ (Abb. 3) interpoliert werden. Hierfür wurden neue, leistungsfähige im Frequenzbereich arbeitende Algorithmen entwickelt. Bei der Optimierung dieser Verfahren musste unter anderem dafür gesorgt werden, dass keine zusätzlichen nichtlinearen Verzerrungen auftraten (24 Bit Signalqualität). Die tatsächlich erreichte Winkelauflösung entspricht der Genauigkeit des Headtrackers (12 Bit). Es stehen Software-Treiber für verschiedene auf dem Markt befindliche Tracking Systeme zur Verfügung, die über eine RS232 Schnittstelle angeschlossen werden können. Über diese können auch in Verbindung mit einem PC neue Raumdaten eingespeist werden. Maximal acht vollständige Räume können derzeit gleichzeitig gespeichert und direkt umgeschaltet werden.

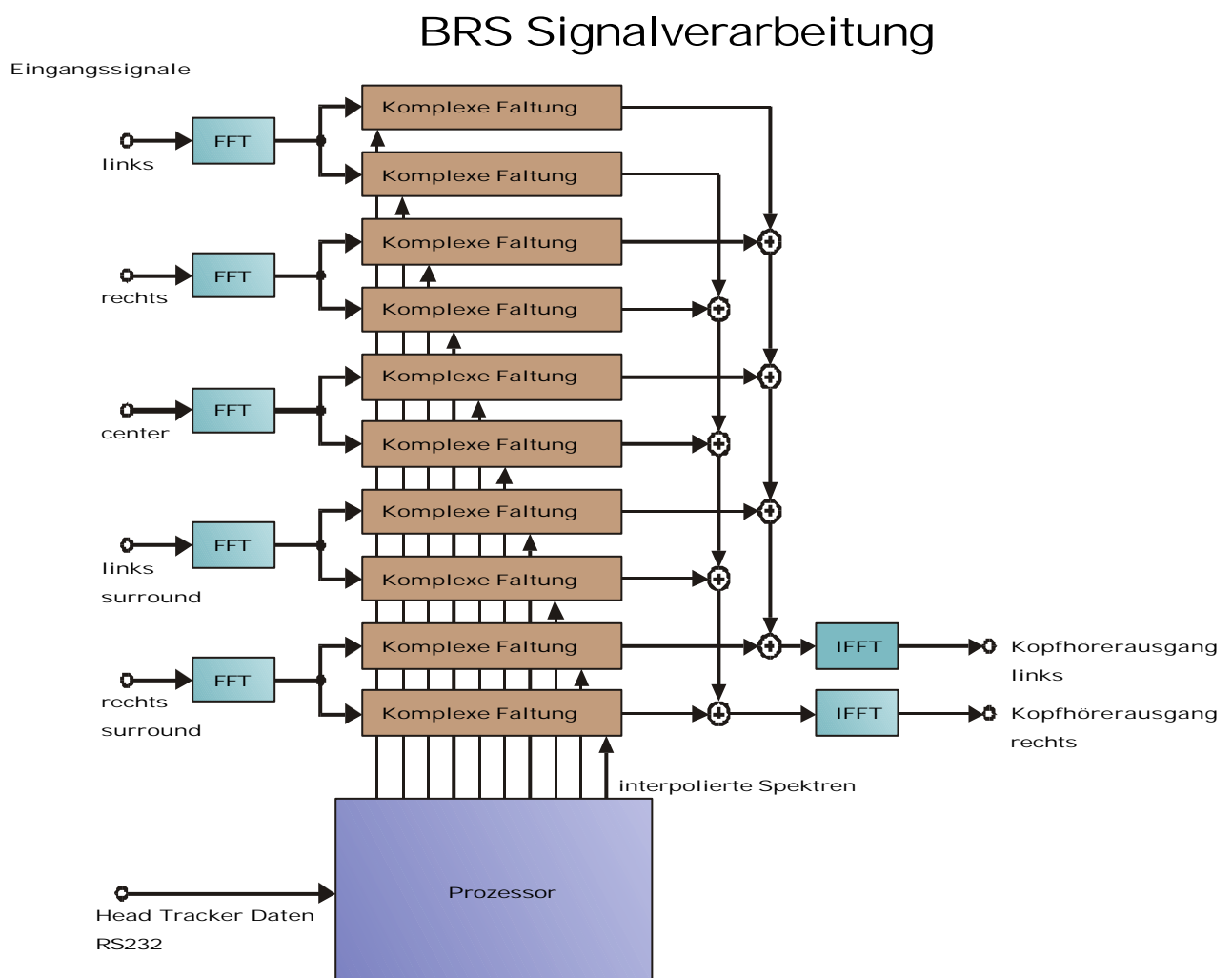


Abb. 3: Signalverarbeitung im BRS Prozessor

Wir wollen noch auf die folgenden Optionen hinweisen:

- Für Multiuserbetrieb ist ein 0°- Vorneabgleich des Headtrackers vorgesehen. Sinnvoll ist dies, wenn z.B. mehrere Benutzer mit Kopfhörern arbeiten und schräg auf einen Bildschirm im Zentrum schauen.
- Die Verwendung hochwertiger, diffusfeldentzerrter Kopfhörer ist unabdingbar für ein gutes Ergebnis. Für ausgewählte Marken stehen Entzerrfilter zur Verfügung, die im Konfigurationsmenu fest anwählbar sind.
- Bekanntlich ist eines der Hauptprobleme bei der Kopfhörerwiedergabe die unnatürliche, „körperlose“ Wiedergabe tiefer Frequenzen. Als durchaus befriedigende Lösung hat sich der Einsatz eines monauralen Subwoofers erwiesen, der die tiefen Frequenzen bis etwa 150Hz extern abstrahlt. Ein Ausgang dafür ist im Gerät integriert.

4 Applikationen

Basierend auf der BRS-Technologie lassen sich ausgereifte Applikationen für virtuelle Räume realisieren. Sie beschränken sich zunächst auf die Nachbildung existierender Räume, dies aber mit dem Vorteil einer exakten Reproduktion der räumlichen Wahrnehmung im Ursprungsraum. Eine erste Anwendung der BRS-Processors ist der virtuelle Surround Sound Abhörraum. Er gibt dem Tonmeister im Prinzip die Möglichkeit, einen hochwertigen Abhörraum einschließlich der gewohnten Lautsprecher „mitzunehmen“, beispielsweise in den Ü-Wagen.

Bereits für heutige Fernseh- und Radioproduktionen ist Mehrkanalton gefragt (Dolby-Surround), für die digitalen Medien (DVD, DVB, DAB) wird diskreter Mehrkanalton zunehmend eingesetzt werden. Das setzt geeignete Abhörbedingungen voraus, die in vielen Fällen nur mit großem finanziellen Aufwand oder nur unzureichend realisierbar sind. Selbst die konventionelle 2-Kanal-Stereofonie erfordert aufwendige raumakustische Maßnahmen und teure Studiolautsprecher, welche sich durch Einsatz eines virtuellen (Referenz-) Abhörtraumes vermeiden lassen. Der virtuelle Abhörraum kann praktisch in jedem ruhigen Raum problemlos sowohl für Zweikanal-Monitoring als auch für Mehrkanal-Monitoring genutzt werden. Auch mit Blick auf neue Formen der Tonproduktion (z.B. „Redakteursarbeitsplatz“, „Home-Recording“) ergeben sich hier beachtliche Vorteile. Der

virtuelle Abhörraum würde zudem die einzig praktikable und gleichzeitig kostengünstige Möglichkeit bieten, unter standardisierten Bedingungen zu produzieren. Dies gilt insbesondere für die Fernseh-Tonbearbeitung.

Neben deutlicher Kostensenkung ergeben sich zusätzliche praktische Vorteile: Per Umschaltung kann der Tonmeister während der Produktion bestimmte Raum- oder Lautsprechereigenschaften sowie die Lautsprecheranordnung verändern und vergleichen. Das ist wichtig für die Beurteilung der Kompatibilität, beispielsweise Studiowiedergabe mit Heimwiedergabe, oder Filmtone im großen Kino mit Fernsehton. Auch besteht die Möglichkeit, spezielle Abhörsituationen vergleichend zu berücksichtigen, z.B. die Wiedergabequalität einer Mehrkanalanlage im Kraftfahrzeug.

Interessant ist darüber hinaus die Anwendung im Monitor-Einspielweg, dergestalt, daß die Solisten im Studio über den Monitor-Kopfhörer keinen verfälschenden Downmix mit Im-Kopf-Lokalisation erhalten, sondern denselben Höreindruck wie der Tonmeister im Regieraum.

Auch für Consumeranwendungen wäre eine kostengünstige Version der virtuellen Surround Sound Lautsprecherwiedergabe eine sinnvolle Zusatzentwicklung, denn die zukünftigen 5.1 Mehrkanal-Geräte sollten einen hochwertigen 5→2 Downmix für Kopfhörer zur Verfügung stellen.

Die Akzeptanz des Kopfhörers läßt sich für die unterschiedlichen Situationen schwer einschätzen. Fest steht, daß ein wesentlicher Nachteil der Kopfhörerwiedergabe beseitigt ist: weder Im-Kopf-Lokalisation noch kopfbezogene Hörereignisse verändern das Hörbild der Lautsprecherwiedergabe, sondern natürliches räumliches Hören ist in einem raumbezogenen virtuellen Schallfeld gewährleistet. Auch sollte der hohe Tragekomfort moderner Kopfhörer dazu beitragen, eine weitgehende Kopfhörer-Akzeptanz herbeizuführen angesichts des Vorteils, hochwertige (virtuelle) Lautsprecher einzusetzen unter optimalen (virtuellen) raumakustischen, überall reproduzierbaren raumakustischen Bedingungen.

5 Ausblick

Die folgenden Erweiterungen des vorgestellten Systems sind geplant:

- Die Option, über ein Lautsprecherpaar im Nahfeld anstelle des Kopfhörers abzuhören. Um hier eine „Quasi-Kopfhörerwiedergabe“ zu gewährleisten, muss der vom jeweiligen Lautsprecher auf das gegenüberliegende Ohr übersprechende Schallanteil mit Hilfe eines sogenannten „Transauralfilters“ (cross talk canceller) kompensiert werden. Bisher vom Markt angebotene nach diesem Prinzip arbeitende Systeme, insbesondere im Consumer-Bereich, sind noch weit davon entfernt die hier gesetzten Qualitätsanforderungen zu erfüllen. Grundsätzlich ist auch hier die Verwendung des Headtrackers unumgänglich, wobei nun auch laterale Bewegungen des Hörers relativ zu den Abhörlautsprechern erkannt und kompensiert werden müssen.
- Erste Erfahrungen mit dem BRS Prozessor haben gezeigt, dass einige Versuchspersonen frontale Hörereignisse zwar vorn, aber etwas zu hoch wahrnehmen. Dieser Elevationseffekt ist dadurch zu erklären, dass die individuelle Pinna (Ohrmuschel) der Testperson zu stark von der des Kunstkopfes abweicht. Hier ist eine „Ein-Knopf“ Kompensation vorgesehen, die nach Gehör mit geeigneten Testsignalen am Gerät vorgenommen werden kann, wobei entsprechende Filter aktiviert werden, mit denen o.g. Unterschiede kompensiert werden.
- Neben ausgemessenen realen Räumen wäre es auch denkbar synthetische, durch eine Raumsimulation erzeugte zu verwenden. Das Simulationsprogramm muss lediglich die binauralen Raumdaten in gleicher Weise wie die gemessenen generieren. Man kann z.B. auf vorhandene HRTF- Datenbanken zurückgreifen, die simulierten Raumechos gemäss ihrer Einfallrichtung mit den passenden HRTFs falten und erhält so die gewünschten Impulsantworten. Eine denkbare Anwendung wäre die Konstruktion eines idealisierten Referenz-Abhörraums, mit ebenso „idealen“ Lautsprechern, zu Vergleichszwecken oder zur Untersuchung neuer Surroundformate unter idealisierten Bedingungen.

Literatur

- [1] Philip Mackensen, Klaus Reichenauer und Günther Theile: Einfluss der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören. Vorliegender Tagungsband.
- [2] Uwe Felderhoff, Philip Mackensen und Günther Theile: Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale. Vorliegender Tagungsband.