

Beispiele für den Audio-Up-/-Downmix und die Cross-Konversion

Auf DVD und Blu-ray Disc sind in der Regel Stereo- und Surround-Tonspuren parallel vorhanden. Der Anwender kann dabei je nach Ausstattung zwischen Stereo- und Surround-Ton auswählen. Auch bei HDTV hat der Zuschauer die Möglichkeit zwischen Stereo- und Surround-Ton-Wiedergabe zu wählen. Das Stereosignal muss dabei in ein Multikanal-Surround-Signal in Echtzeit konvertiert werden. Generell sollte aber auch eine Cross-Konvertierung möglich sein. Im Beitrag werden anhand von Beispielen unterschiedliche Lösungen aufgezeigt.

On DVD and Blu-ray Disc, stereo and surround-sound tracks exist in parallel so users can select between them. Viewers of HDTV broadcasts also have the option of choosing between stereo or surround-sound so the stereo signal has to be converted into a multi-channel format or vice-versa in realtime. This article shows, by use of examples, different ways to achieve this.

Einleitung

Große hochauflösende Bildschirme sorgen zu Hause beim Zuschauer für ein visuelles Kinoerlebnis. Da eine „umgebende“ Klangwiedergabe das Erlebnis komplettiert, steigt die Popularität von Surround-Sound-Anlagen im Heimbereich. Allerdings sind diese Geräte oft im oberen Preissegment angesiedelt. Gleichzeitig erwarten die Anwender eine ideale Wiedergabe der Inhalte und geben sich nicht mit Stereo und schon gar nicht mit wechselnden Formaten zufrieden.

Auf DVD und Blu-ray Disc sind in der Regel Stereo- und Surround-Tonspuren nebeneinander vorhanden. Der Konsument kann dabei je nach Ausstattung zwischen Stereo- und Surround-Sound auswählen. Auch beim Fernsehen werden seit der Einführung von DVB mehrere Tonspuren parallel zum Bild ausgestrahlt, daher steigert die Einführung von HDTV die Popularität von Surround-Sound auch fürs Fernsehen. Der Zuschauer hat hier ebenfalls die Möglichkeit zwischen Stereo- und Surround-Sound-Wiedergabe auszuwählen. Das setzt voraus, dass stets Surround-Sound zur Verfügung steht, ansonsten müsste das System alternativlos auf Stereo zurückschalten.

Ist ein Film nur in Stereo aufgezeichnet oder ein Live-Programm wird nur in Stereo produziert, bedarf es, damit eine kontinuierliche Tonübertragung sichergestellt ist, einer Konvertierung der Stereo-Information in ein Multikanal-Surround-Signal in Echtzeit. In der Vergangenheit wurden solche Produktionen im Tonstudio nachbearbeitet.

Diese Nachbearbeitungen sind zum Teil schwierig, da oft die ursprüngliche Aufnahmesituation nicht bekannt ist. Somit sind einige

Versuche nötig, um eine optimale Einstellung zu finden. Das ist ein aufwendiger und zeitraubender Prozess. Geräte, die in Echtzeit arbeiten, vereinfachen die Vorbereitung und die Umsetzung erheblich. Dazu kommt noch, dass die Tonmodellierung bei einer TV-Produktion nicht standardisiert ist und Genre-bezogen spezifische Anforderungen haben.

Das wichtigste Ziel ist, dem Zuschauer die „Situation“ in einer Weise zu präsentieren, dass er das Gefühl hat, dabei zu sein. Der Aufwand für die Realisierung ist am Aufnahmeort schon für Stereoproduktionen erheblich. Bei Surround-Aufnahmen ist er um einiges höher. Darüber hinaus sind bei der Modellierung der Signale ganz neue Aspekte zu berücksichtigen, wie zum Beispiel bei Live-Sportübertragungen.

Für jede Sportart sind spezielle Anforderungen zu erfüllen. Die einzige Konstante bei diesen Aufnahmen ist die Position des Moderators vorne im Center. Der Sprecher hat bei der Sendung oder Aufnahme in der Regel sein eigenes Mikrofon. Es ist letztendlich unerheblich, ob der Moderator nur aus dem dafür vorgesehenen Lautsprecher bei einer Surround-Wiedergabe ertönt oder als Phantomquelle mittig zwischen den Stereolautsprecherboxen.

Grundsätzlich sind alle akustischen Inhalte von einem Ursprung stammend als Stereo- und als 5.1-Signal identisch. Nur die Verteilung ist, bedingt durch die differierende Anzahl der Kanäle, unterschiedlich.

Unter Berücksichtigung dieser These sollte es auch möglich sein, ein Surround-Signal in ein Stereo-Signal zu wandeln und umgekehrt: Von Surround auf Stereo sind mithilfe einer Downmix-Matrix sehr gute Ergebnisse zu erzielen.

Beim Konvertieren von Stereo in Surround

verhält es sich etwas schwieriger, da die Informationen aus zwei Kanälen auf fünf Kanäle zu verteilen sind. In professionellen Systemen ist es unabdingbar, dass derartige Geräte systemfähig sind und in der Produktionskette keine zusätzlichen Artefakte einbringen. Das könnte zu unerwünschten Effekten führen, nicht nur in einer Kaskade unterschiedlicher Prozesse innerhalb der Übertragungsstrecke, auch beim Zuschauer kann das, bedingt durch Inkompatibilitäten, negative Auswirkungen haben.

Systeme

Die einfachste Methode zum Konvertieren von Stereo- in Surround-Material ist die Nutzung passiver oder auch dynamischer Matrixschaltungen. Während der Vorteil dieser Systeme die kurze Latenzzeit ist, ist die unbefriedigende Zuordnung der Signale von Nachteil. Diese Systeme sind nur bedingt für hochwertige Produktionen geeignet.

Erheblich bessere Ergebnisse erzielt man durch die spektrale Analyse der Druckgradienten und die Zuordnung von Direkt- und Raumschallanteilen. Der Vorteil dieser Methoden ist eine wesentlich bessere Zuordnung der Schallanteile. Der Nachteil ist die wesentlich längere Latenzzeit. Derartige Ansätze sind in Hardware-Systemen integriert, in denen der Algorithmus durch leistungsfähige Signalprozessoren in Echtzeit abgearbeitet wird. Aufgrund dieser Fähigkeit können sie in Live-Produktionen eingesetzt werden. Weiterhin gibt es auch reine PC-Software-Lösungen, die für den Einsatz in Postproduktionen gedacht sind.

Isostem

Das Isostem-System (1 HE, vier AES-Ein-/Ausgänge) arbeitet nach dem Verfahren der spektralen Analyse der Druckgradienten. Nachfolgend werden die beiden Varianten vorgestellt:

- Die Variante „Expert“ verfügt über alle Einstellmöglichkeiten. Mit diesem System lassen sich alle Parameter über ein GUI, ähnlich wie bei einem Mischpult, vom Anwender einstellen. Der Toningenieur ist dadurch in der Lage, genau seine Abstimmung vorzunehmen und damit seine eigenen Presets zu gestalten. Bis zu sechs Presets können in den Geräten gespeichert

Peter Petersen (FKTG)
ist Berater im Bereich
Audiotechnik



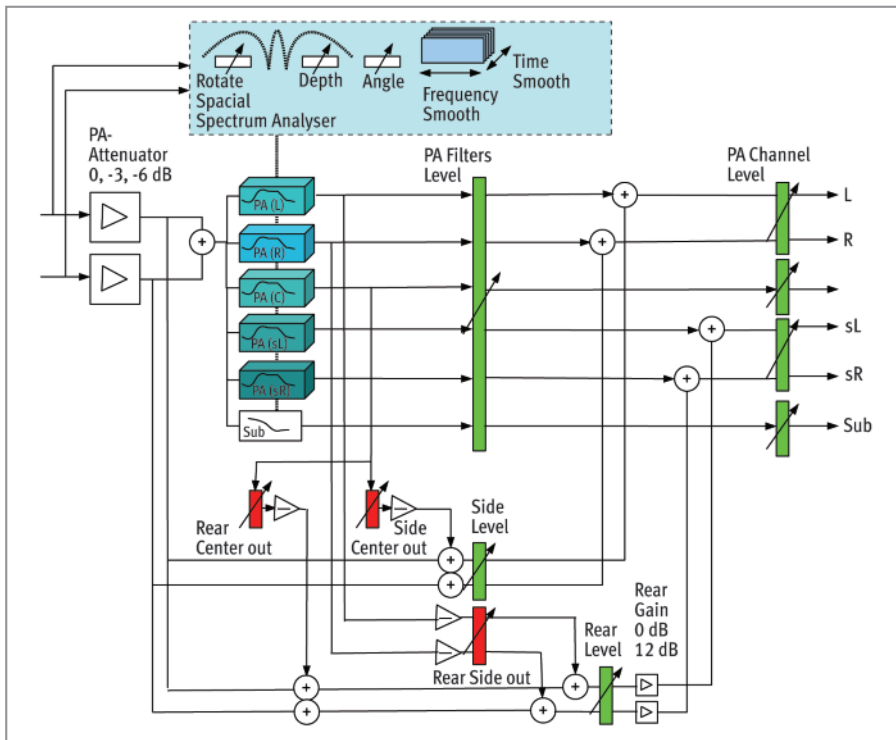


Bild 1. Blockschaltbild des Panoramic-Analyzers

- und abgerufen werden.
 - Bei der „Live“-Variante besteht nicht die Möglichkeit, eigene Presets mittels eines GUIs selbst zu erstellen. Stattdessen sind hier feste Presets werksseitig voreingestellt. Die Presets können allerdings durch andere, zum Beispiel durch die vom Toningenieur optimierten, ersetzt werden. Beide Varianten verfügen über maximal sechs Preset-Speicher, die über Schalteingänge aktivierbar sind.
- Die Signalverarbeitung und somit die Konvertierungsergebnisse sind in beiden Systemen identisch.

Das „Live“-System ist für den Einsatz im Sendebetrieb gedacht. Es verfügt zusätzlich über eine automatische Havarie-Funktion: Zwei Geräte sind koppelbar und schalten beim Ausfall eines Gerätes automatisch um.

Grundsätzliche Idee

Der angewandte Algorithmus basiert auf dem Prinzip der spektralen Analyse und wurde von Antoine Hurtado, Honorarprofessor am National Conservatory of Paris, erfunden. Er hat in einem Vortrag auf der AES 2005 in New York das Prinzip der Öffentlichkeit vorgestellt, nachdem er es zum Patent angemeldet hatte (s. Audio Engineering Society: AES Paper 6548; AES Convention 119; October 2005).

Die Grundidee des Algorithmus ist, dass dominante Quellen, insbesondere Sprache und Musikinstrumente, durch eine spektrale intensimetrische Analyse identifizierbar sind. Durch diese Feststellung wurde eine Lösung für eine fundamentale Fragestellung gefun-

den: Was ist bei Multikanalmaterial Direkt-schall und was ist Umgebungsschall?

Festlegung der Quellen im Umfeld „Panoramic Analyzer“

Der „Panoramic Analyzer“ im Isostem-System ermöglicht die Ermittlung der räumlichen Energieverteilung eines Stereo-Programms. Genau gesagt, unterscheidet das Verfahren zwischen einem „trockenen“ Signal, also Direkt-schall, und einem Umgebungssignal mit Echoanteilen. Zusätzlich bietet es in einem zentralen Fenster des GUIs einen Überblick über die Verteilung der Intensitätsvektoren. All diese Informationen geben dem Toningenieur die Möglichkeit, einen Upmix mit allen relevanten räumlichen Quellen zu erzeugen und einen Surround-Raum zu schaffen. Diese Berechnungen sind reversibel, da sie keine zusätzlichen Informationen hinzufügen, sondern nur die vorhandenen Informationen selektieren und neu verteilen. Am Ausgang des Prozesses stehen diskrete Kanäle zur Verfügung, die als Multitrack aufgezeichnet oder auch zu einem Dolby-Encoder geführt werden können.

Bild 1 zeigt das Blockschaltbild des Panoramic-Analyzers. Oben sieht man den „akustischen Spektrum Analysator“. Dieser untersucht das Stereosignal auf den Ursprung der spektralen Druckgradienten, ermittelt den Energiefluss, steuert die darunter befindlichen Filterbänke und reproduziert das akustische Modell auf die Surround-Kanäle. Mit den Parametern „Rotate“, „Depth“ und „Angle“ justiert man die Mitte, den Öffnungswinkel und die Fil-

tergüte. Die Parameter „Frequency-Smooth“ und „Time Smooth“ werden zur Optimierung der Filterbänke eingesetzt.

In den nachfolgenden Stufen werden durch gezielte Addition und Subtraktion die Kanäle angepasst und unerwünschte Anteile, wie „der Center-Kanal in den Seitenkanälen“ oder „die Seitenkanäle und der Center-Kanal in den Surround-Kanälen“ unterdrückt. In ihrer Gesamtheit sind diese Einstellungsmöglichkeiten die Basis für eine akzentuierte Bestimmung der Selektivität. Die Zuordnung der akustischen Informationen zu den jeweiligen Lautsprechern ist bedingt durch die feinen Abstimmungsmöglichkeiten für nahezu alle Genres gut geeignet.

„Virtual Microphone“

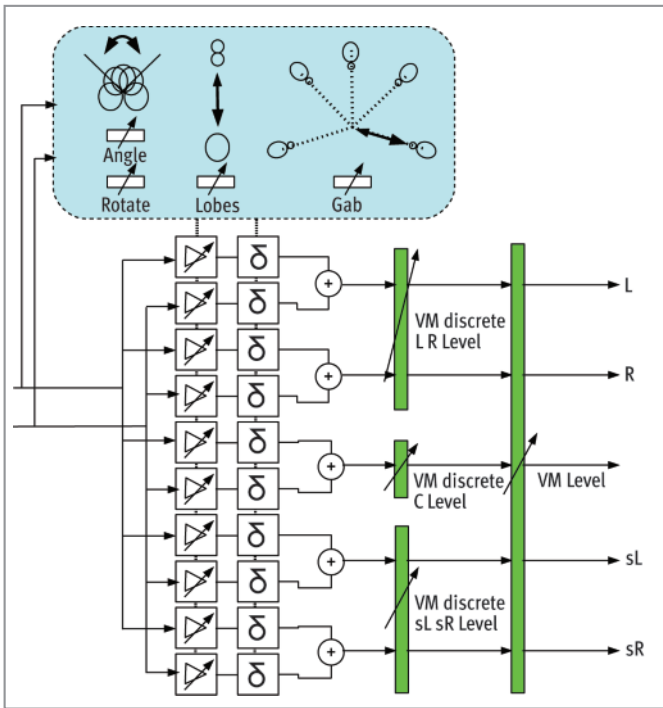
Das zusätzlich integrierte Verfahren „Virtual Microphone“ basiert, wie der Panoramic-Analyser, auf einer akustischen Simulation der Stereo-Quelle. Dabei wird die Stereo-Information auf ein virtuelles Mikrofonsystem übertragen (**Bild 2**). Es lassen sich die Parameter Öffnungswinkel („Angle“), Richtcharakteristik („Lobes“) und Mikrofonabstand („Gap“) im Analysator einstellen. Das berechnete Modell wird durch die darunterliegenden, vom Analyser gesteuerten Pegelsteller und Zeitglieder auf die Ausgangskanäle projiziert. Zur Feinabstimmung für eigene Anforderungen können noch die Pegel korrigiert werden. Im „Virtual Microphone“-Modul wird der Sub-Kanal, ähnlich wie auch bei einer realen Surround-Mikrofonierung, nicht berücksichtigt.

Beide Verfahren arbeiten vollkommen autonom und können parallel zum Einsatz kommen. Die Vorteile beider Verfahren lassen sich somit verbinden und problematische Aufnahmen mit guten Ergebnissen konvertieren.

Weitere Funktionen

Der Panoramic-Analyser und das Virtual-Microphone verarbeiten die Signale im internen Bussystem im entsprechenden Frequenzbereich. Anschließend erfolgt eine Phasen- und Frequenzanpassung in den seitlichen und rückwärtigen Kanälen (**Bild 3**). Die „Phase“-Funktion gleicht die Phasenlage der seitlichen und hinteren Kanäle aus. Eine Up-Konversion ruft eine negative Tendenz bezüglich der L-R- und sL-sR-Korrelation hervor.

Das Ursprungssignal und der Center haben einen festen Korrelationsbezug zueinander. Durch die Konversion verändert sich die Phasenlage zum Center sowie den Seiten- und hinteren Kanälen. Die Phase-Funktion korrigiert diesen Effekt, indem sie die Phase der L-



R- und sL-sR-Signale variabel einstellbar verschiebt. Die Winkelbezüge der Seiten- und Rückkanäle sind entsprechend der Lautsprecheranordnung kalibriert und können nur gemeinsam variiert werden.

In der Regel enthalten die hinteren Kanäle nicht die gleichen Frequenzanteile, wie die vorderen Kanäle. Mit dem HCF-Filter besteht die Möglichkeit die hinteren Kanäle zu korrigieren, damit ein korrektes Signal an den hinteren Kanälen ausgegeben wird.

Alle genannten Verfahren bieten dem Toningenieur vielfältige Möglichkeiten, um ein Stereo- zu einem Surround-Signal zu modellieren. Für die Gestaltung neuer Konfigurationen kann man vorhandene Presets einsetzen und diese durch gezielte Feinabstimmung optimieren.

ISO-Funktion und Downmix

Schlussendlich besteht die Möglichkeit, Mehrkanalsignale direkt mit dem „Upmix“ zu mischen und zum Ausgang weiterzuleiten oder weiteren Stufen für „Downmix“ sowie „Intermix“ anzubieten. Im System (Bild 4 unten links) befinden sich zwei „Downmix“-Stufen. Eine wird für interne Zwecke im Intermix genutzt. Die zweite „Downmix“-Stufe hat einen Ausgang und kann parallel zum „Upmix“ verwendet werden.

Eine weitere Funktion ist der Vergleich eines Stereo-Referenzsignals mit dem internen „Downmix“. Oben links im Bild 4 befindet sich die interne „Downmix“-Matrix, deren Eingänge die Signale von der Mischstufe darunter erhalten. Die Differenz beider Signale (Stereo-Referenz und „Downmix“) wird zur sogenannten „ISO-Stufe“ weitergeleitet. Im Idealfall ist

am „Intermix“-Ausgang kein Signal vorhanden, da beide Eingangssignale theoretisch identisch sind. Die ISO-Funktion gewährleistet die Kompatibilität zwischen Stereo- und Multikanalformaten.

Die Differenz zwischen dem „Downmix“- und dem Stereosignal (der sogenannte Intermix) wird in das Multikanalsignal injiziert. Zur besseren Intermix-Integration im Multikanal-Datenstrom werden die Signale in „MS“ konvertiert. Die ISO-Stufe separiert die Elemente „M“ und „S“ des Intermix-Signals. Das Element „M“ wird innerhalb der Kanäle „C“ und „L-R“, das Element „S“ wird zwischen „L-R“ und „sL-sR“ aufgeteilt.

Die „MS“-Signale sind einstellbar und

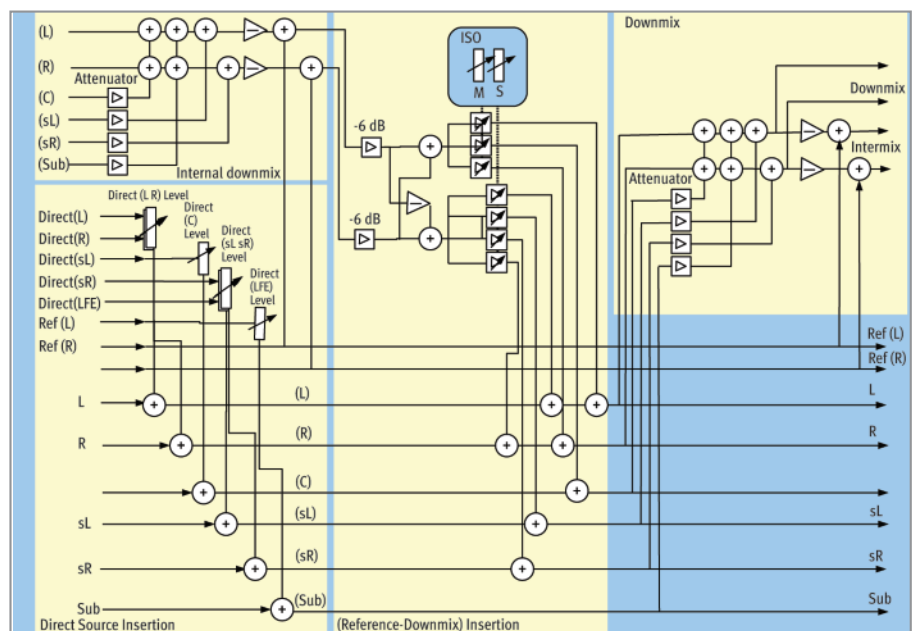


Bild 4. Isystem-Systemübersicht

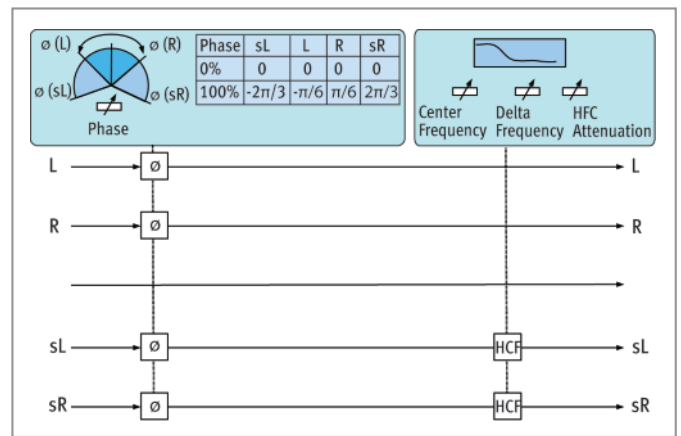


Bild 3. Phasen- und Frequenzanpassung in den seitlichen und rückwärtigen Kanälen

Bild 2. Übertragung der Stereo-Information auf ein virtuelles Mikrofonsystem

erlauben eine freie Zuordnung von „M“ oder „S“, entweder zum Center oder zu den seitlichen Kanälen.

Da alle Funktionen parallel arbeiten und gleichzeitig zur Verfügung stehen, können sie beliebig miteinander kombiniert werden und stellen dem Anwender einen Werkzeugkasten mit vielen Bearbeitungsmöglichkeiten zur Verfügung.

Workflow-Beispiele

Bild 5 zeigt einen „Crossmix“, bei dem zwei Stereosignale anliegen. Das eine mit einem Kommentatorsignal und das andere nur mit dem Raumsignal (Atmo). Das Stereo-„Main“-Signalpaar durchläuft die Upmix-Stufen und danach die Phasenkorrektur als 5.1-Format. Beide Signale durchlaufen das „Intermix“- und „ISO“-Modul. Ziel ist es den Kommentator vollständig aus den Seitenkanälen und den hinteren Kanälen herauszufiltern, damit er nur im

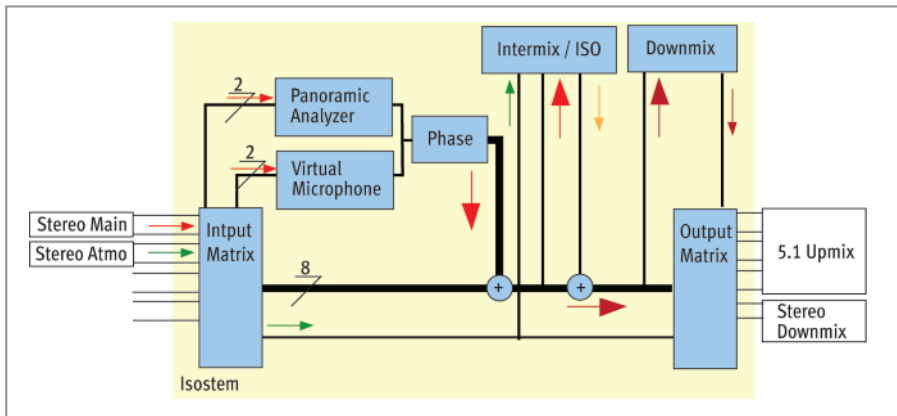


Bild 5. Konfiguration eines Crossmixes mit zwei Stereo-Signalen
←

Center erscheint. **Bild 6** zeigt ein Upmix mit nur einer Stereoquelle. Hier verhält es sich ähnlich, wie im zuvor beschriebenen Fall. Es gibt aber kein zusätzliches Signal für die Intermix- und ISO-Funktionen, die damit nicht eingesetzt werden.

Die bisherigen Beispiele zeigen nur „Upmix“-Szenarien. Das System kann natürlich auch 5.1-Signale durchleiten bzw. daraus zusätzlich einen „Downmix“ bilden, der parallel ausgegeben wird (**Bild 7**).

In einem Programmablauf kann es vorkommen, dass Programmteile mit unterschiedlichen Formaten zur Ausspielung kommen, wie zum Beispiel sequentiell 5.1 und Stereo. Das würde in herkömmlichen Systemen unweigerlich zu Problemen führen, da am Ausgang nicht ein einheitliches Format angeboten werden könnte. Gerade bei zeitkritischen Ereignissen ist ein System mit einer derartigen Funktion vorteilhaft, da sich eine vorausgehende Bearbeitung erübrigt. Das bedingt ein hohes Maß an Qualität und Presets, damit sie bei der Wiedergabe ein konstantes Verhalten haben.

Über sechs Presetspeicher, die über Schalteingänge umschaltbar sind, lassen sich unterschiedlichste Konfigurationen aufrufen. Verbindet man das System mit einer Sendeaufautomation, kann diese für das jeweilige Sound-Format den voreingestellten Preset setzen und am Ausgang konstant ein 5.1-Signal mit dem dazugehörigen Downmix bereitstellen (**Bild 8**).

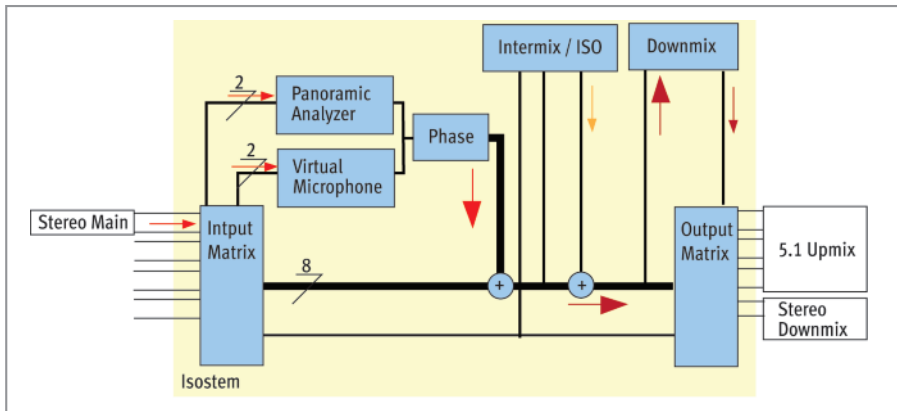


Bild 6. Typisches Setup für einen Upmix

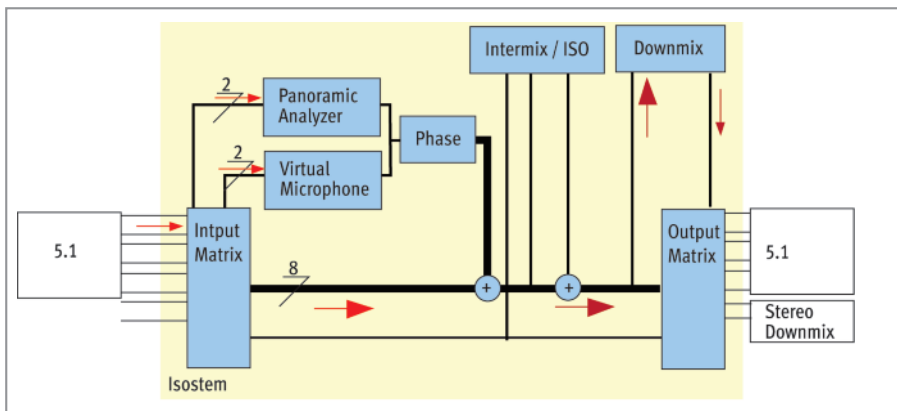


Bild 7. Downmix-Setup zur Erzeugung eines 5.1- und Stereo-Signals

Schlussbemerkung

Die Popularität von HD-Bild- und Surround-Tonformaten wird in den nächsten Jahren durch den steigenden Absatz von Flachbildschirmen und Surround-Anlagen zunehmen. In den Archiven gibt es aber viele Produktionen mit Stereoton. Somit gibt es einen massiven Bedarf nach Lösungen zur Umsetzung von Stereo- auf Surround-Ton.

Darüber hinaus ist der Einsatz von echtzeitfähigen Geräten bei Live-Sendungen eine logische Konsequenz, denn damit erreicht man Flexibilität bei der Produktion: Unterschiedliche Produktionsorte, die verschiedene Sound-Formate übertragen, können somit, ohne dass der Zuschauer den Unterschied wahrnimmt, direkt in Folge gesendet werden.

Nach Unterlagen von Isostem (User Guide v.1201) und einem Artikel in der Sonovision (Isostem, et le son coule de source).

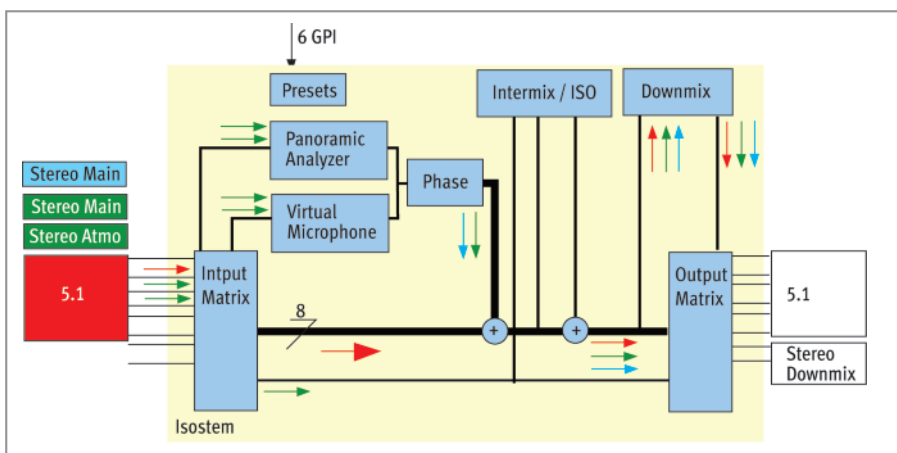


Bild 8. Downmix aus Stereo- und 5.1-Signalen