



(11)

EP 2 094 031 A2

(12)

## EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:  
26.08.2009 Patentblatt 2009/35

(51) Int Cl.:  
**H04S 3/00 (2006.01)**

**G10L 19/00 (2006.01)**

(21) Anmeldenummer: 09006142.5

(22) Anmeldetag: 22.02.2006

(84) Benannte Vertragsstaaten:

**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR  
HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI  
SK TR**

(30) Priorität: 04.03.2005 DE 102005010057

(62) Dokumentnummer(n) der früheren Anmeldung(en)  
nach Art. 76 EPÜ:  
**06707184.5 / 1 854 334**

(71) Anmelder: **Fraunhofer-Gesellschaft zur  
Förderung der  
angewandten Forschung e.V.  
Hansastraße 27c  
80636 München (DE)**

(72) Erfinder:

- **Plogsties, Jan**  
91052 Erlangen (DE)
- **Mundt, Harald**  
91058 Erlangen (DE)
- **Popp, Harald**  
90587 Tuchenbach (DE)

(74) Vertreter: **Zinkler, Franz et al**  
**P.O. Box 246**  
**82043 Pullach (DE)**

### Bemerkungen:

Diese Anmeldung ist am 05-05-2009 als  
Teilanmeldung zu der unter INID-Code 62 erwähnten  
Anmeldung eingereicht worden.

(54) **Vorrichtung und Verfahren zum Erzeugen eines codierten Stereo-Signals eines Audioträgers oder Audiodatenstroms**

(57) Eine Vorrichtung zum Erzeugen eines codierten Stereo-Signals aus einer Multikanal-Darstellung umfasst einen Multikanal-Decodierer (11), der aus wenigstens einem Basiskanal und Parameterinformationen drei oder mehr Multi-Kanäle erzeugt. Die drei oder mehr Multi-Kanäle werden Kopfhörer-Signalverarbeitung (12) unterzogen, um einen uncodierten ersten Stereo-Kanal und einen uncodierten zweiten Stereo-Kanal zu erzeugen, wel-

che dann einem Stereo-Codierer (13) zugeführt werden, um ausgangsseitig eine codierte Stereo-Datei zu erzeugen. Die codierte Stereo-Datei kann jedem geeigneten Wiedergabegerät in Form eines CD-Players oder eines Hardware-Players zugeführt werden, sodass ein Benutzer des Wiedergabegeräts nicht nur einen normalen Stereo-Eindruck erhält, sondern einen Multi-Kanal-Eindruck bekommt.

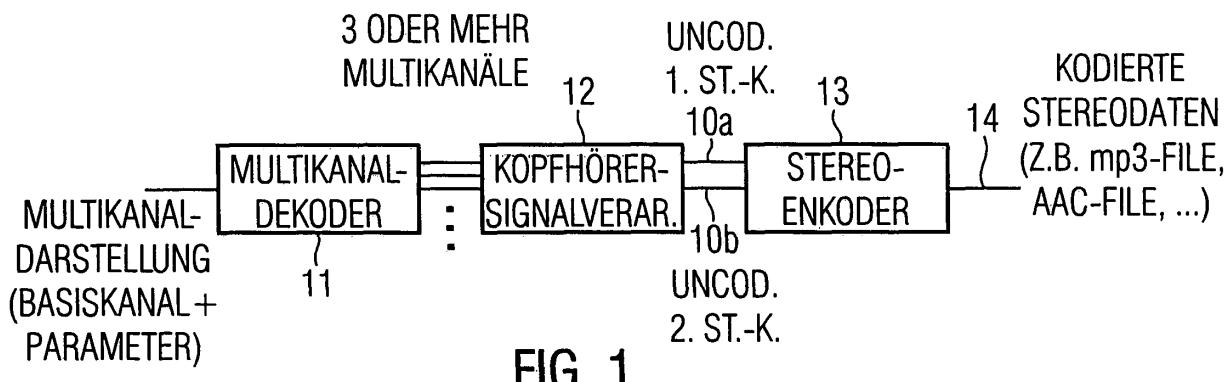


FIG. 1

## Beschreibung

**[0001]** Die vorliegende Erfindung bezieht sich auf die Multikanal-Audioteknik und insbesondere auf Multikanal-Audioanwendungen in Verbindung mit Kopfhörer-Techniken.

**[0002]** Die beiden internationalen Patentanmeldungen WO 99/49574 und WO 99/14983 offenbaren Audiosignalverarbeitungstechniken zum Ansteuern eines Paares von gegenüberliegend angeordneten Kopfhörer-Lautsprechern, damit ein Benutzer über die beiden Kopfhörer eine räumliche Wahrnehmung der Audioszene erhält, die nicht nur eine Stereo-Darstellung sondern eine Multikanal-Darstellung ist. So erhält der Hörer über seine bzw. ihre Kopfhörer eine räumliche Wahrnehmung eines Audiotracks, die im besten Fall gleich seiner räumlichen Wahrnehmung ist, wenn der Benutzer in einem Wiedergaberaum sitzen würde, der beispielsweise mit einer 5.1- Audioanlage ausgestattet ist. Zu diesem Zweck wird für jeden Kopfhörer-Lautsprecher jeder Kanal des Multikanal-Audiotracks oder Multikanal-Audiodatenstroms, wie es in Fig. 2 dargestellt ist, einem eigenen Filter zugeführt, wonach dann die jeweils gefilterten zusammengehörenden Kanäle aufaddiert werden, wie es nachfolgend dargestellt wird.

**[0003]** Auf einer linken Seite in Fig. 2 befinden sich die Multikanal-Eingänge 20, die zusammen eine Multikanal-Darstellung des Audiotracks oder Audiodatenstroms repräsentieren. Ein solches Szenario ist beispielsweise in Fig. 10 schematisch gezeigt. Fig. 10 zeigt einen Wiedergaberaum 200, in dem eine so genannte 5.1-Audioanlage angeordnet ist. Die 5.1-Audioanlage umfasst einen Mitte-Lautsprecher 201, einen Vorne-Links-Lautsprecher 202, einen Vorne-Rechts-Lautsprecher 203, einen Hinten-Links-Lautsprecher 204 und einen Hinten-Rechts-Lautsprecher 205. Eine 5.1-Audioanlage hat einen zusätzlichen Subwoofer 206, der auch als Low-Frequency-Enhancement-Kanal bezeichnet wird. Im so genannten "Sweet Spot" des Wiedergaberaums 200 befindet sich ein Zuhörer 207, der einen Kopfhörer 208 trägt, welcher einen linken Kopfhörer-Lautsprecher 209 und einen rechten Kopfhörer-Lautsprecher 210 aufweist.

**[0004]** Die in Fig. 2 gezeigte Verarbeitungseinrichtung ist nunmehr ausgebildet, um jeden Kanal 1, 2, 3 der Multikanal-Eingänge 20 mit einem Filter  $H_{1L}$ , der den Schallkanal vom Lautsprecher zum linken Lautsprecher 209 in Fig. 10 beschreibt, zu filtern, und um denselben Kanal ferner mit einem Filter  $H_{1R}$  zu filtern, der den Schall von einem der fünf Lautsprecher zum rechten Ohr bzw. zum rechten Lautsprecher 210 des Kopfhörers 208 darstellt.

**[0005]** Wäre beispielsweise der Kanal 1 in Fig. 2 der vordere linke Kanal, der durch den Lautsprecher 202 in Fig. 10 ausgestrahlt wird, so würde das Filter  $H_{1L}$  den durch eine gestrichelte Linie 212 angedeuteten Kanal darstellen, während das Filter  $H_{1R}$  den durch eine gestrichelte Linie 213 dargestellten Kanal wiedergeben würde. Wie es in Fig. 10 beispielsweise durch eine gestrichelte Linie 214 angedeutet ist, erhält der linke Kopfhörerlaut-

sprecher 209 nicht nur den Direktschall, sondern auch frühe Reflexionen an einer Rand des Wiedergaberaums und natürlich auch späte Reflexionen, die in einem diffusen Nachhall ausgedrückt werden.

**[0006]** Eine solche Filterdarstellung ist in Fig. 11 dargestellt. Insbesondere zeigt Fig. 11 ein schematisches Beispiel für eine Impulsantwort eines Filters, beispielsweise des Filters  $H_{1L}$  von Fig. 2 dar. Der Direktschall, der durch die Linie 212 in Fig. 11 dargestellt ist, wird durch einen Peak am Anfang des Filters dargestellt, während frühe Reflexionen, wie sie beispielsweise durch 214 in Fig. 10 dargestellt sind, durch einen mittleren Bereich mit mehreren (diskreten) kleineren Peaks in Fig. 11 wiedergegeben werden. Der diffuse Nachhall ist dann typischerweise nicht mehr nach einzelnen Peaks aufgelöst, da der Schall des Lautsprechers 202 prinzipiell beliebig oft reflektiert wird, wobei die Energie natürlich mit jeder Reflexion und zusätzlicher Ausbreitungsstrecke weiter abnimmt, wie es durch die abnehmende Energie im hinteren Abschnitt, der mit "diffuser Nachhall" in Fig. 11 bezeichnet ist, dargestellt ist.

**[0007]** Jedes der in Fig. 2 gezeigten Filter umfasst daher eine Filter-Impulsantwort, die in etwa einen Verlauf hat, wie er durch die schematische Impulsantwortdarstellung in Fig. 11 wiedergegeben ist. Selbstverständlich wird die einzelne Filter-Impulsantwort vom Wiedergaberaum, der Positionierung der Lautsprecher, eventueller Dämpfungseigenschaften im Wiedergaberaum z. B. aufgrund mehrerer anwesender Personen oder im Wiedergaberaum befindlichen Möbeln etc. sowie idealerweise auch von den Eigenschaften der einzelnen Lautsprecher 201 bis 206 abhängen.

**[0008]** Die Tatsache, dass sich die Signale von allen Lautsprechern am Ohr des Zuhörers 207 superponieren, wird durch die Addierer 22 und 23 in Fig. 2 dargestellt. Es wird also jeder Kanal mit einem entsprechenden Filter für das linke Ohr gefiltert, um dann die von den Filtern ausgegebenen Signale, die alle für das linke Ohr bestimmt sind, einfach aufzuhaddieren, um das Kopfhörer-Ausgangssignal für das linke Ohr L zu erhalten. Analog wird eine Addition durch den Addierer 23 für das rechte Ohr bzw. für den rechten Kopfhörer-Lautsprecher 210 in Fig. 10 vorgenommen, um durch Überlagerung sämtlicher mit einem entsprechenden Filter für das rechte Ohr gefilterten Lautsprecher-Signale das Kopfhörer-Ausgangssignal für das rechte Ohr zu erhalten.

**[0009]** Aufgrund der Tatsache, dass es neben dem Direktschall auch frühe Reflexionen und insbesondere auch einen diffusen Nachhall gibt, welche insbesondere für die Raumwahrnehmung von großer Bedeutung sind, damit der Ton nicht synthetisch oder "hölzern" klingt, sondern dem Hörer das Gefühl vermittelt, er sitzt wirklich in einem Konzertsaal mit seinen akustischen Eigenschaften, werden die Impulsantworten der einzelnen Filter 21 alle eine beträchtliche Länge annehmen. Die Faltung jedes einzelnen Multi-Kanals der Multikanal-Darstellung mit zwei Filtern führt daher bereits zu einer erheblichen Rechenaufgabe. Da für jeden einzelnen Multi-Kanal zwei

Filter benötigt werden, nämlich einer für das linke Ohr und ein anderer für das rechte Ohr, werden, wenn der Subwoofer-Kanal ebenfalls eigens behandelt wird, für eine Kopfhörer-Wiedergabe einer 5.1-Multikanal-Darstellung insgesamt 12 voneinander unterschiedliche Filter benötigt. Alle Filter haben, wie es aus Fig. 11 ersichtlich ist, eine sehr lange Impulsantwort, um nicht nur den Direktschall sondern auch frühe Reflexionen und den diffusen Nachhall berücksichtigen zu können, der einem Audiotrack eigentlich erst die richtige Klangwiedergabe und einen guten Raumeindruck verleiht.

**[0010]** Um das bekannte Konzept in die Realität umzusetzen, wird daher, wie es in Fig. 10 gezeigt ist, neben einem Multikanalspieler 220 die sehr aufwendige virtuelle Tonverarbeitung 222 benötigt, die die Signale für die beiden Lautsprecher 209 und 210, die durch Leitungen 224 und 226 in Fig. 10 dargestellt sind, liefert.

**[0011]** Kopfhörer-Anlagen zum Erzeugen eines Multikanal-Kopfhörer-Sounds sind daher aufwendig, sperrig und teuer, was an der hohen Rechenleistung, am hohen Strombedarf für die nötige hohe Rechenleistung und am hohen Arbeitsspeicherbedarf für die durchzuführenden Bewertungen mit der Impulsantwort und am damit verbundenen großen Volumen oder teuren Bausteinen für das Abspielgerät liegt. Solche Anwendungen sind daher an Home-PC-Soundkarten bzw. Laptop-Soundkarten oder auch Heim-Stereo-Anlagen gebunden.

**[0012]** Insbesondere ist dem immer größer werdenden Markt an mobilen Abspielgeräten, wie beispielsweise mobilen CD-Playern oder insbesondere den Hardware-Playern bleibt der Multikanal-Kopfhörer-Sound verschlossen, da die Rechenanforderungen zum Filtern der Multi-Kanäle mit z. B. 12 unterschiedlichen Filtern sowohl im Hinblick auf die Prozessorressourcen als auch im Hinblick auf den Stromverbrauch der typischerweise batteriebetriebenen Geräte nicht in dem Preissegment realisierbar sind. Hier geht es um ein Preissegment am unteren (niedrigen) Ende der Skala. Gerade dieses Preissegment ist aber aufgrund der großen Stückzahlen wirtschaftlich sehr interessant.

**[0013]** Die Aufgabe der vorliegenden Erfindung besteht darin, ein effizientes Signalverarbeitungskonzept zu liefern, das eine Kopfhörer-Wiedergabe in Multikanal-Qualität auch auf einfachen Wiedergabegeräten ermöglicht.

**[0014]** Diese Aufgabe wird durch eine Vorrichtung zum Erzeugen eines codierten Stereo-Signals nach Patentanspruch 1 oder durch ein Verfahren zum Erzeugen eines codierten Stereo-Signals nach Patentanspruch 11 oder ein Computer-Programm nach Patentanspruch 12 gelöst.

**[0015]** Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, dass der hoch qualitative und attraktive Multikanal-Kopfhörer-Sound allen verfügbaren Abspielgeräten, wie beispielsweise CD-Playern oder Hardware-Playern, dadurch zur Verfügung gestellt werden kann, dass aus einer Multikanal-Darstellung eines Audiotracks oder Audiodatenstroms, also beispielsweise einer 5.1-Dar-

stellung eines Audiotracks außerhalb eines Hardware-Players, also z. B. in einem rechenstarken Computer eines Providers einer Kopfhörer-Signalverarbeitung unterzogen. Erfindungsgemäß wird das Ergebnis Kopfhörer-Signalverarbeitung jedoch nicht einfach abgespielt, sondern einem typischen Audio-Stereo-Codierer zugeführt, der dann aus dem linken Kopfhörerkanal und dem rechten Kopfhörerkanal ein codiertes Stereosignal erzeugt.

**[0016]** Dieses codierte Stereosignal kann dann, wie jedes andere codierte Stereosignal, das keine Multikanaldarstellung aufweist, dem Hardware-Player oder z. B. einem mobilen CD-Player in Form einer CD zugeführt werden. Das Wiedergabegerät wird dann den Benutzer mit einem Kopfhörer-Multikanal-Sound versorgen, ohne dass irgendwelche zusätzlichen Ressourcen bzw. Einrichtungen an bereits bestehenden Geräten hinzugefügt werden müssen. Erfindungsgemäß wird das Ergebnis der Kopfhörer-Signalverarbeitung, also das linke und das rechte Kopfhörersignal nicht, wie im Stand der Technik, in einem Kopfhörer wiedergegeben, sondern codiert und als codierte Stereo-Daten ausgegeben.

**[0017]** Eine solche Ausgabe kann eine Speicherung, eine Übertragung oder irgendetwas ähnliches sein. Eine solche Datei mit codierten Stereodaten kann dann ohne weiteres jedem beliebigen für Stereo-Wiedergabe ausgebildeten Wiedergabegerät zugeführt werden, ohne dass der Benutzer irgendwelche Änderungen an seinem Gerät durchführen müsste.

**[0018]** Das erfindungsgemäße Konzept, aus dem Ergebnis der Kopfhörer-Signalverarbeitung ein codiertes Stereo-Signal zu erzeugen, ermöglicht es somit, dass die Multikanal-Darstellung, die eine wesentlich bessere und lebensnahere Qualität für einen Benutzer liefert, auch auf allen einfachen und sehr stark verbreiteten und in Zukunft noch stärker verbreiteten Hardware-Playern eingesetzt werden kann.

**[0019]** Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird von einer codierten Multikanal-Darstellung ausgegangen, also einer Parameterdarstellung, die einen oder typischerweise zwei Basiskanäle aufweist, und die ferner Parameterdaten aufweist, um auf der Basis der Basiskanäle und der Parameterdaten die Multi-Kanäle der Multikanaldarstellung zu erzeugen. Nachdem ein Frequenzbereichbasiertes Verfahren zur Multikanal-Decodierung bevorzugt wird, wird erfindungsgemäß die Kopfhörer-Signalverarbeitung nicht in dem Zeitbereich durch Faltung des Zeitsignals mit der Impulsantwort durchgeführt, sondern im Frequenzbereich durch Multiplikation mit der Filter-Übertragungsfunktion.

**[0020]** Dies ermöglicht die Einsparung von wenigstens einer Rücktransformation vor der Kopfhörer-Signalverarbeitung und ist insbesondere dann vorteilhaft, wenn auch der nachfolgende Stereo-Codierer im Frequenzbereich arbeitet, sodass dann, ohne dass jemals in den Zeitbereich gegangen werden muss, die Stereo-Codierung des Kopfhörer-Stereo-Signals ebenfalls ohne Gang in den Zeitbereich erfolgen kann. Die Verarbeitung von

der Multikanal-Darstellung bis zum codierten Stereosignal ohne Einschaltung eines Zeitbereichs oder durch eine wenigstens reduzierte Anzahl von Transformationen ist nicht nur im Hinblick auf die Rechenzeiteffizienz interessant, sondern grenzt Qualitätsverluste ein, da weniger Verarbeitungsstufen auch weniger Artefakte in das Audiosignal einführen.

**[0021]** Insbesondere bei Block-basierten Verfahren, die unter Berücksichtigung einer psychoakustischen Maskierungsschwelle quantisieren, wie es für den Stereo-Codierer bevorzugt wird, ist es wichtig, so viel Tandem-Codierungs-Artefakte als möglich zu verhindern.

**[0022]** Bei einem besonders bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird als Multikanal-Darstellung eine BCC-Darstellung mit einem oder vorzugsweise zwei Basiskanälen verwendet. Nachdem das BCC-Verfahren im Frequenzbereich arbeitet, werden die Multi-Kanäle nicht, wie beim BCC-Decodierer üblich, nach ihrer Synthese in den Zeitbereich transformiert. Stattdessen wird die blockweise vorliegende Spektraldarstellung der Multikanäle verwendet und der Kopfhörer-Signalverarbeitung unterzogen. Hierzu werden die Übertragungsfunktionen der Filter verwendet, also die Fourier-Transformierten der Impulsantworten, um eine Multiplikation zwischen der Spektraldarstellung der Multi-Kanäle und den Filter-Übertragungsfunktionen durchzuführen. Sofern die Impulsantworten der Filter zeitlich länger als ein Block von Spektralkomponenten am Ausgang des BCC-Decodierers sind, wird eine blockweise Filterverarbeitung bevorzugt, bei der die Impulsantworten der Filter im Zeitbereich getrennt werden und blockweise transformiert werden, um dann entsprechende für solche Maßnahmen nötige Gewichtungen der Spektren durchzuführen, wie sie beispielsweise in der WO 94/01933 offenbart sind.

**[0023]** Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend Bezug nehmend auf die beiliegenden Zeichnungen detailliert erläutert. Es zeigen:

Fig. 1 ein Blockschaltbild der erfindungsgemäßen Vorrichtung zum Erzeugen eines codierten Stereo-Signals;

Fig. 2 eine Detaildarstellung einer Implementierung der Kopfhörer-Signalverarbeitung von Fig. 1;

Fig. 3 einen bekannten Joint-Stereo-Codierer zum Erzeugen von Kanaldaten und parametrischen Multikanal-Informationen;

Fig. 4 eine Darstellung eines Schemas zum Bestimmen von ICLD-, ICTD- und ICC-Parametern für eine BCC-Codierung/Decodierung;

Fig. 5 eine Blockdiagrammdarstellung einer BCC-Codierer/Decodierer-Kette;

Fig. 6 ein Blockdiagramm einer Implementierung des BCC-Synthese-Blocks von Fig. 5;

Fig. 7 eine Kaskadierung zwischen einem Multikanal-Decodierer und der Kopfhörer-Signalverarbeitung ohne Transformation in den Zeitbereich;

Fig. 8 eine Kaskadierung zwischen der Kopfhörer-Signalverarbeitung und einem Stereo-Codierer ohne Transformation in den Zeitbereich;

Fig. 9 ein Prinzip-Blockdiagramm eines bevorzugten Stereo-Codierers;

Fig. 10 eine Prinzipdarstellung eines Wiedergabe-Szenarios zum Bestimmen der Filterfunktionen von Fig. 2;

Fig. 11 eine prinzipielle Darstellung einer zu erwartenden Impulsantwort eines Filters, das gemäß Fig. 10 bestimmt ist.

**[0024]** Fig. 1 zeigt ein Prinzip-Blockschaltbild einer erfindungsgemäßen Vorrichtung zum Erzeugen eines codierten Stereo-Signals eines Audioträcks oder Audiodatenstroms. Das Stereo-Signal umfasst in uncodierter Form einen uncodierten ersten Stereo-Kanal 10a sowie eine uncodierten zweiten Stereo-Kanal 10b und wird aus einer Multikanal-Darstellung des Audioträcks oder Audiodatenstroms erzeugt, wobei die Multikanal-Darstellung Informationen über mehr als zwei Multi-Kanäle aufweist. Wie später noch dargestellt wird, kann die Multikanal-Darstellung in einer uncodierten oder codierten Form vorliegen. Ist die Multikanal-Darstellung in uncodierter Form vorhanden, so umfasst sie drei oder mehr Multi-Kanäle. Bei einem bevorzugten Anwendungs-Szenario umfasst die Multikanal-Darstellung fünf Kanäle und einen Subwoofer-Kanal.

**[0025]** Ist die Multikanal-Darstellung dagegen in einer codierten Form vorhanden, so umfasst diese codierte Form typischerweise einen oder mehrere Basis-Kanäle sowie Parameter zum Synthesieren der drei oder mehr Multi-Kanäle aus dem einen oder den beiden Basiskanälen. Ein Multikanal-Decodierer 11 ist daher ein Beispiel für eine Einrichtung zum Bereitstellen der mehr als zwei Multi-Kanäle aus der Multikanal-Darstellung. Liegt die Multikanal-Darstellung dagegen bereits in uncodierter Form vor, also z. B. in Form von 5+1 PCM-Kanälen, so entspricht die Einrichtung zum Bereitstellen einem Eingangsanschluss für eine Einrichtung 12 zum Durchführen einer Kopfhörer-Signalverarbeitung, um das uncodierte Stereosignal mit dem uncodierten ersten Stereo-Kanal 10a und dem uncodierten zweiten Stereo-Kanal 10b zu erzeugen.

**[0026]** Vorzugsweise ist die Einrichtung 12 zum Durchführen Kopfhörer-Signalverarbeitung ausgebildet, um die Multi-Kanäle der Multikanal-Darstellung jeweils

mit einer ersten Filterfunktion für den ersten Stereo-Kanal und mit einer zweiten Filterfunktion für den zweiten Stereo-Kanal zu bewerten und bewertete Multi-Kanäle jeweils aufzuaddieren, um den uncodierten ersten Stereo-Kanal und den uncodierten zweiten Stereo-Kanal zu erhalten, wie es anhand von Fig. 2 dargestellt worden ist. Der Einrichtung 12 zum Durchführen der Kopfhörer-Signalverarbeitung ist ein Stereo-Codierer 13 nachgeschaltet, der ausgebildet ist, um den ersten uncodierten Stereo-Kanal 10a und den zweiten uncodierten Stereo-Kanal 10b zu codieren, um das codierte Stereo-Signal an einem Ausgang 14 des Stereo-Codierers 13 zu erhalten. Der Stereo-Codierer führt eine Datenraten-Reduktion durch, sodass eine Datenrate, die zum Übertragen des codierten Stereo-Signals nötig ist, kleiner als eine Datenrate ist, die zum Übertragen des uncodierten Stereo-Signals nötig ist.

**[0027]** Erfundungsgemäß wird somit ein Konzept erreicht, das es ermöglicht, Mehrkanalton, der auch als "Surround" bezeichnet wird, über einfache Abspielgeräte, wie beispielsweise Hardware-Player, Stereo-Kopfhörern zuzuführen.

**[0028]** Als einfache Kopfhörer-Signalverarbeitung kann z. B. die Summe bestimmter Kanäle gebildet werden, um die Ausgangskanäle für die Stereo-Daten zu erhalten. Verbesserte Verfahren arbeiten mit komplexeren Algorithmen, die wiederum eine bessere Qualität der Wiedergabe erreichen.

**[0029]** Es sei darauf hingewiesen, dass es das erfundungsgemäß Konzept ermöglicht, dass die rechenaufwendigen Schritte zum Multikanal-Decodieren und zum Durchführen der Kopfhörer-Signalverarbeitung nicht im Abspielgerät selbst durchgeführt werden, sondern extern durchgeführt werden. Das Ergebnis des erfundungsgemäß Konzepts ist eine codierte Stereo-Datei, die beispielsweise ein MP3-File, ein AAC-File, ein HE-AAC-File oder irgendein anderes Stereo-File ist.

**[0030]** Bei anderen Ausführungsbeispielen können auch die Multikanal-Decodierung, die Kopfhörer-Signalverarbeitung und die Stereo-Codierung auf unterschiedlichen Geräten ausgeführt werden, da die Ausgangsdaten bzw. Eingangsdaten der einzelnen Blöcke leicht portierbar und standardisiert erzeugbar und abspeicherbar sind.

**[0031]** Nachfolgend wird Bezug nehmend auf Fig. 7 eine bevorzugte Ausführungsform der vorliegenden Erfahrung dargestellt, bei der der Multikanal-Decodierer 11 eine Filterbank oder eine FFT-Funktion aufweist, derart, dass die Multikanal-Darstellung im Frequenzbereich geliefert wird. Im Einzelnen werden die einzelnen Multi-Kanäle als Blöcke von Spektralwerten für jeden Kanal separat erzeugt. Erfundungsgemäß wird dann die Kopfhörer-Signalverarbeitung nicht im Zeitbereich durch Faltung der zeitlichen Kanäle mit den Filter-Impulsantworten durchgeführt, sondern es wird eine Multiplikation der Frequenzbereichs-Darstellung der Multikanäle mit einer Spektraldarstellung der Filter-Impulsantwort durchgeführt. Am Ausgang der Kopfhörer-Signalverarbeitung

wird dann ein uncodiertes Stereosignal erreicht, das jedoch nicht im Zeitbereich vorliegt, sondern das einen linken und einen rechten Stereo-Kanal umfasst, wobei ein solcher Stereo-Kanal als Folge von Blöcken von Spektralwerten gegeben ist, wobei jeder Block von Spektralwerten ein Kurzzeitspektrum des Stereo-Kanals darstellt.

**[0032]** Bei dem in Fig. 8 gezeigten Ausführungsbeispiel wird der Kopfhörer-Signalverarbeitungs-Block 12 eingangsseitig entweder mit Zeitbereichs- oder Frequenzbereichsdaten versorgt. Ausgangsseitig werden die uncodierten Stereo-Kanäle im Frequenzbereich, also wieder als Folge von Blöcken von Spektralwerten erzeugt. Als Stereo-Codierer 13 wird in diesem Fall ein Stereo-Codierer bevorzugt, der Transformations-basiert ist, der also Spektralwerte verarbeitet, ohne dass zwischen der Kopfhörer-Signalverarbeitung 12 und dem Stereo-Codierer 13 eine Frequenz/Zeit-Umsetzung und eine anschließende Zeit-Frequenz-Umsetzung erforderlich ist. Ausgangsseitig gibt der Stereo-Codierer 13 dann eine Datei mit dem codierten Stereosignal aus, die neben Seiteninformationen eine codierte Form von Spektralwerten umfasst.

**[0033]** Bei einem besonders bevorzugten Ausführungsbeispiel der vorliegenden Erfahrung wird auf dem Weg von der Multikanal-Darstellung am Eingang des Blocks 11 von Fig. 1 bis zur codierten Stereo-Datei am Ausgang 14 der Einrichtung von Fig. 1 eine durchgehende Frequenzbereichs-Verarbeitung durchgeführt, ohne dass eine Umsetzung in den Zeitbereich und gegebenenfalls wieder eine Umsetzung in den Frequenzbereich zu erfolgen hat. Wird als Stereo-Codierer ein MP3-Codierer oder ein AAC-Codierer eingesetzt, so wird bevorzugt, das Fourier-Spektrum am Ausgang des Kopfhörer-Signalverarbeitungs-Blocks in ein MDCT-Spektrum umzusetzen. Damit wird erfundungsgemäß sichergestellt, dass die Phaseninformationen, die für die Faltung/Bewertung der Kanäle im Kopfhörer-Signalverarbeitungs-Block exakt benötigt werden, in die nicht derart phasenkorrekt arbeitende MDCT-Darstellung umgerechnet werden, sodass für den Stereo-Codierer im Gegensatz zu einem normalen MP3-Codierer oder einem normalen AAC-Codierer keine Einrichtung zum Umsetzen von Zeitbereich in den Frequenzbereich, also in das MDCT-Spektrum benötigt wird.

**[0034]** Fig. 9 zeigt ein allgemeines Blockschaltbild für einen bevorzugten Stereo-Codierer. Der Stereo-Codierer umfasst eingangsseitig ein Joint-Stereo-Modul 15, das vorzugsweise adaptiv bestimmt, ob eine gemeinsame Stereocodierung beispielsweise in Form einer Mitte/Seite-Codierung einen höheren Codiergewinn liefert als eine getrennte Verarbeitung von linkem und rechtem Kanal. Das Joint-Stereo-Modul 15 kann ferner ausgebildet sein, um eine Intensity-Stereo-Codierung durchzuführen, wobei eine Intensity-Stereo-Codierung insbesondere bei höheren Frequenzen einen beträchtlichen Codiergewinn liefert, ohne dass hörbare Artefakte auftreten. Der Ausgang des Joint-Stereo-Moduls 15 wird dann un-

ter Verwendung verschiedener weiterer Redundanz-reduzierender Maßnahmen, wie beispielsweise einer TNS-Filterung, einer Rauschsubstitution etc. weiterverarbeitet, um dann die Ergebnisse einem Quantisierer 16 zuzuführen, der unter Verwendung einer psychoakustischen Maskierungsschwelle eine Quantisierung der Spektralwerte erreicht. Die Quantisierer-Schrittweite ist dabei derart gewählt, dass das durch das Quantisieren eingeführte Rauschen unterhalb der psychoakustischen Markierungsschwelle bleibt, sodass eine Datenratenreduktion erreicht wird, ohne dass die durch die verlustbehaftete Quantisierung eingeführten Verzerrungen hörbar werden. Dem Quantisierer 16 ist schließlich ein Entropie-Codierer 17 nachgeschaltet, der eine verlustlose Entropie-Codierung der quantisierten Spektralwerte durchführt. Am Ausgang des Entropie-Codierers liegt dann das codierte Stereosignal vor, das neben den Entropie-codierten Spektralwerten zur Decodierung nötige Seiteninformationen umfasst.

**[0035]** Nachfolgend wird auf bevorzugte Implementierungen des Multikanal-Decodierers bzw. auf bevorzugte Multikanal-Darstellungen anhand der Fig. 3 bis 6 eingegangen.

**[0036]** So existieren in der Technik viele Techniken zum Reduzieren der Datenmenge, die zur Übertragung eines Multikanal-Audiosignals benötigt wird. Solche Techniken werden Joint-Stereo-Techniken genannt. Zu diesem Zweck wird auf Fig. 3 verwiesen, die eine Joint-Stereo-Vorrichtung 60 zeigt. Diese Vorrichtung kann eine Vorrichtung sein, die beispielsweise die Intensity-Stereo- (IS-) Technik oder die Binaural Cue Codieretechnik (BCC) implementiert. Ein solches Gerät empfängt üblicherweise als Eingangssignal zumindest zwei Kanäle CH1, CH2, ..., CHn, und gibt einen einzigen Trägerkanal sowie parametrische Multikanalinformationen aus. Die parametrischen Daten sind so definiert, dass in einem Decodierer eine Approximation eines Ursprungskanals (CH1, CH2, ..., CHn) berechnet werden kann.

**[0037]** Normalerweise wird der Trägerkanal Subband-Abtastwerte, Spektralkoeffizienten, Zeitbereichsabtastwerte etc. umfassen, die eine relativ feine Darstellung des zugrundeliegenden Signals liefern, während die parametrischen Daten keine solchen Abtastwerte oder Spektralkoeffizienten umfassen, sondern Steuerparameter zum Steuern eines bestimmten Rekonstruktionsalgorithmus, wie beispielsweise Gewichten durch Multiplizieren, durch Zeitverschieben, durch Frequenzverschieben, etc. Die parametrischen Multikanalinformationen umfassen daher eine relativ grobe Darstellung des Signals oder des zugeordneten Kanals. In Zahlen ausgedrückt beträgt die Menge an Daten, die von einem Trägerkanal benötigt wird, eine Menge von etwa 60 bis 70 kBit/s, während die Menge an Daten, die durch parametrische Seiteninformationen für einen Kanal benötigt wird, im Bereich von 1,5 bis 2,5 kBit/s ist. Es sei darauf hingewiesen, dass die vorstehenden Zahlen für komprimierte Daten gelten. Selbstverständlich benötigt ein nicht-komprimierter CD-Kanal Datenraten im Bereich

von etwa dem Zehnfachen. Ein Beispiel für parametrische Daten sind die bekannten Skalenfaktoren, Intensity-Stereo-Informationen oder BCC-Parameter, wie es nachfolgend dargelegt wird.

**[0038]** Die Technik der Intensity-Stereo-Codierung ist in dem AES-Preprint 3799, "Intensity Stereo Coding", J. Herre, K.H. Brandenburg, D. Lederer, Februar 1994, Amsterdam beschrieben. Allgemein basiert das Konzept von Intensity Stereo auf einer Hauptachsentransformation, die auf Daten beider stereophoner Audiokanäle durchzuführen ist. Wenn die meisten Datenpunkte um die erste Hauptachse herum konzentriert sind, kann ein Codiergewinn erreicht werden, indem beide Signale um einen bestimmten Winkel gedreht werden, bevor die Codierung stattfindet. Dies ist jedoch nicht immer für reale stereophone Reproduktionstechniken gegeben. Daher wird diese Technik dahingehend modifiziert, dass die zweite orthogonale Komponente von der Übertragung in dem Bitstrom ausgeschlossen wird. Somit bestehen die rekonstruierten Signale für den linken und den rechten Kanal aus unterschiedlich gewichteten oder skalierten Versionen desselben übertragenen Signals. Dennoch unterscheiden sich die rekonstruierten Signale in ihrer Amplitude, sie sind jedoch identisch im Hinblick auf ihre Phaseninformationen. Die Energie-Zeit-Hüllkurven beider ursprünglicher Audiokanäle werden jedoch durch die selektive Skalierungsoperation beibehalten, die typischerweise auf frequenzselektive Art und Weise arbeitet. Dies entspricht der menschlichen Wahrnehmung des Schalls bei hohen Frequenzen, wo die dominanten räumlichen Informationen durch die Energiehüllkurven bestimmt werden.

**[0039]** Zusätzlich wird bei praktischen Implementierungen das übertragene Signal, d. h. der Trägerkanal aus dem Summensignal des linken Kanals und des rechten Kanals anstatt der Rotation beider Komponenten erzeugt. Ferner wird diese Verarbeitung, d. h. das Erzeugen von Intensity-Stereo-Parametern zum Durchführen der Skalierungsoperationen frequenzselektiv durchgeführt, d. h. unabhängig für jedes Skalenfaktorband, d. h. für jede Codiererfrequenzpartition. Vorzugsweise werden beide Kanäle kombiniert, um einen kombinierten oder "Träger"-Kanal und zusätzlich zu dem kombinierten Kanal die Intensity-Stereo-Informationen zu bilden. Die Intensity-Stereo-Informationen hängen von der Energie des ersten Kanals, der Energie des zweiten Kanals oder der Energie des kombinierten Kanals ab.

**[0040]** Die BCC-Technik ist in dem AES-Convention-Paper 5574 "Binaural Cue Coding applied to stereo and multichannel audio compression", T. Faller, F. Baumgarde, Mai 2002, München, beschrieben. Bei der BCC-Codierung wird eine Anzahl von Audioeingangskanälen in eine Spektraldarstellung umgewandelt, und zwar unter Verwendung einer DFT-basierten Transformation mit überlappenden Fenstern. Das resultierende Spektrum wird in nicht-überlappende Abschnitte eingeteilt, von denen jeder einen Index hat. Jede Partition hat eine Bandbreite proportional zu der äquivalenten Rechteck-

bandbreite (ERB). Die Inter-Kanal-Pegelunterschiede (ICLD; ICLD = Inter Channel Level Differences) und die Interkanal-Zeitunterschiede (ICTD; ICTD = Inter Channel Time Differences) werden für jede Partition und für jeden Frame k ermittelt. Die ICLD und ICTD werden quantisiert und codiert, um schließlich als Seiteninformationen in einen BCC-Bitstrom zu kommen. Die Interkanal-Pegelunterschiede und die Interkanal-Zeitunterschiede sind für jeden Kanal relativ zu einem Referenzkanal gegeben. Dann werden die Parameter gemäß vorbestimmter Formeln berechnet, die von den bestimmten Partitionen des zu verarbeitenden Signals abhängen.

**[0041]** Auf Decodiererseite empfängt der Decodierer typischerweise ein Monosignal und den BCC-Bitstrom. Das Monosignal wird in den Frequenzbereich transformiert und in einen Raumsyntheseblock (Spatial-Syntheseblock) eingegeben, der auch decodierte ICLD- und ICTD-Werte empfängt. In dem Spatial-Syntheseblock werden die BCC-Parameter (ICLD und ICTD) verwendet, um eine Gewichtungsoperation des Monosignals durchzuführen, um die Multikanalsignale zu synthetisieren, die, nach einer Frequenz-/Zeit-Umwandlung eine Rekonstruktion des ursprünglichen Multikanal-Audiosignals darstellen.

**[0042]** Im Fall von BCC ist das Joint-Stereo-Modul 60 wirksam, um die kanalseitigen Informationen so auszugeben, dass die parametrischen Kanaldaten quantisierte und codierte ICLD- oder ICTD-Parameter sind, wobei einer der ursprünglichen Kanäle als Referenzkanal zum Codieren der Kanalseiteninformationen verwendet wird.

**[0043]** Normalerweise wird der Trägersignal aus der Summe der teilnehmenden Ursprungskanäle gebildet.

**[0044]** Natürlich liefern die obigen Techniken nur eine Monodarstellung für einen Decodierer, der nur den Trägerkanal verarbeiten kann, der jedoch nicht in der Lage ist, die parametrischen Daten zur Erzeugung von einer oder mehreren Approximationen von mehr als einem Eingangskanal zu verarbeiten.

**[0045]** Die BCC-Technik ist auch in den US-Patentveröffentlichungen US 2003/0219130 A1, US 2003/0026441 A1 und US 2003/0035553 A1 beschrieben. Zusätzlich wird auf die Fachveröffentlichung "Binaural Cue Coding. Part II: Schemes and Applications", T. Faller und F. Baumgarte, IEEE Trans. On Audio and Speech Proc. Bd. 11, Nr. 6, November 2003 verwiesen.

**[0046]** Nachfolgend wird ein typisches BCC-Schema zur Multikanalaudiocodierung detaillierter dargestellt, und zwar Bezug nehmend auf die Fig. 4 bis 6.

**[0047]** Fig. 5 zeigt ein solches BCC-Schema zur Codierung/Übertragung von Multikanalaudiosignalen. Das Multikanalaudioeingangssignal an einem Eingang 110 eines BCC-Codierers 112 wird in einem sogenannten Downmix-Block 114 heruntergemischt. Bei diesem Beispiel ist das ursprüngliche Multikanalsignal an dem Eingang 110 ein 5-Kanal-Surround-Signal mit einem vorderen linken Kanal, einem vorderen rechten Kanal, einem linken Surround-Kanal, einem rechten Surround-Kanal und einem Mittenkanal. Bei dem bevorzugten Ausfüh-

rungsbeispiel der vorliegenden Erfindung erzeugt der Downmix-Block 114 ein Summensignal durch eine einfache Addition dieser fünf Kanäle in ein Monosignal.

**[0048]** Andere Downmixing-Schemen sind in der 5 Technik bekannt, so dass unter Verwendung eines Multikanal-Eingangssignals ein Downmix-Kanal mit einem einzigen Kanal erhalten wird.

**[0049]** Dieser einzige Kanal wird an einer Summensignalleitung 115 ausgegeben. Eine Seiteninformation, die 10 von dem BCC-Analyseblock 116 erhalten wird, wird auf einer Seiteninformationsleitung 117 ausgegeben.

**[0050]** Bei dem BCC-Analyseblock werden Interkanal-Pegelunterschiede (ICLD) und Interkanal-Zeitunterschiede (ICTD) berechnet, wie es vorstehend dargestellt 15 worden ist. Neuerdings ist der BCC-Analyseblock 116 auch in der Lage, Interkanal-Korrelationswerte (ICC-Werte) zu berechnen. Das Summensignal und die Seiteninformationen werden in einem quantisierten und codierten Format zu einem BCC-Decodierer 120 übertragen. Der BCC-Decodierer zerlegt das übertragene Summensignal in eine Anzahl von Subbändern und führt Skalierungen, Verzögerungen und andere Verarbeitungsschritte aus, um die Subbänder der auszugebenden Multikanal-Audiokanäle zu liefern. Diese Verarbeitung wird 20 so durchgeführt, dass die ICLD-, ICTD- und ICC-Parameter (Cues) eines rekonstruierten Multikanalsignals am Ausgang 121 mit den entsprechenden Cues für das 25 ursprüngliche Multikanalsignal am Eingang 110 in dem BCC-Codierer 112 übereinstimmen. Zu diesem Zweck umfasst der BCC-Decodierer 120 einen BCC-Syntheseblock 122 und einen Seiteninformationenüberarbeitungsblock 123.

**[0051]** Nachfolgend wird der interne Aufbau des BCC-Syntheseblocks 122 Bezug nehmend auf Fig. 6 dargestellt. Das Summensignal auf der Leitung 115 wird in eine Zeit-/Frequenz-Umwandlungseinheit oder Filterbank FB 125 eingespeist. Am Ausgang des Blocks 125 existiert eine Anzahl N von Subbandsignalen oder, in einem Extremfall, ein Block von Spektralkoeffizienten, wenn die 35 Audio-Filterbank 125 eine 1:1-Transformation durchführt, d. h. eine Transformation, die N Spektralkoeffizienten aus N Zeitbereichsabtastwerten erzeugt.

**[0052]** Der BCC-Syntheseblock 122 umfasst ferner eine Verzögerungsstufe 126, eine Pegelmodifikationsstufe 127, eine Korrelationsverarbeitungsstufe 128 und eine Inversfilterbankstufe IFB 129. Am Ausgang der Stufe 129 kann das rekonstruierte Multikanalaudiosignal mit beispielsweise fünf Kanälen im Falle eines 5-Kanal-Surroundsystems zu einem Satz von Lautsprechern 124 40 ausgegeben werden, wie sie in Fig. 5 oder Fig. 4 dargestellt sind.

**[0053]** Das Eingangssignal  $s_n$  wird in den Frequenzbereich oder den Filterbankbereich mittels des Elements 125 umgewandelt. Das Signal, das vom Element 125 45 ausgegeben wird, wird derart kopiert, dass mehrere Versionen desselben Signals erhalten werden, wie es durch den Kopierknoten 130 dargestellt ist. Die Anzahl der Versionen des ursprünglichen Signals ist gleich der Anzahl 50

der Ausgangskanäle in dem Ausgangssignal. Dann wird jede Version des ursprünglichen Signals am Knoten 130 einer bestimmten Verzögerung  $d_1, d_2, \dots, d_i, \dots d_N$  unterzogen. Die Verzögerungsparameter werden durch den Seiteninformationsverarbeitungsblock 123 in Fig. 5 berechnet und von den Interkanal-Zeitunterschieden, wie sie durch den BCC-Analyseblock 116 von Fig. 5 berechnet worden sind, abgeleitet.

**[0054]** Dasselbe gilt für die Multiplikationsparameter  $a_1, a_2, \dots, a_i, \dots, a_N$ , die ebenfalls durch den Seiteninformationsverarbeitungsblock 123 basierend auf den Interkanal-Pegelunterschieden, wie sie durch den BCC-Analyseblock 116 berechnet worden sind, berechnet werden.

**[0055]** Die durch den BCC-Analyseblock 116 berechneten ICC-Parameter werden zum Steuern der Funktionalität des Blocks 128 verwendet, so dass bestimmte Korrelationen zwischen den verzögerten und in ihren Pegeln manipulierten Signalen an den Ausgängen des Blocks 128 erhalten werden. Es sei hier darauf hingewiesen, dass die Reihenfolge der Stufen 126, 127, 128 von der in Fig. 6 gezeigten Reihenfolge abweichen kann.

**[0056]** Es sei darauf hingewiesen, dass bei einer rahmenweisen Verarbeitung des Audiosignals auch die BCC-Analyse rahmenweise durchgeführt wird, also zeitlich variabel, und dass ferner eine frequenzweise BCC-Analyse erhalten wird, wie es durch die Filterbank-Aufteilung aus Fig. 6 ersichtlich ist. Dies bedeutet, dass die BCC-Parameter für jedes Spektralband erhalten werden. Dies bedeutet ferner, dass in dem Fall, in dem die Audiofilterbank 125 das Eingangssignal in beispielsweise 32 Bandpasssignale zerlegt, der BCC-Analyseblock einen Satz von BCC-Parametern für jedes der 32 Bänder erhält. Natürlich führt der BCC-Syntheseblock 122 von Fig. 5, der detailliert in Fig. 6 dargestellt ist, eine Rekonstruktion durch, die auch auf den beispielhaft genannten 32 Bändern basiert.

**[0057]** Nachfolgend wird Bezug nehmend auf Fig. 4 ein Szenario dargestellt, das dazu verwendet wird, um einzelne BCC-Parameter zu bestimmen. Normalerweise können die ICLD-, ICTD- und ICC-Parameter zwischen Kanalpaaren definiert werden. Es wird jedoch bevorzugt, die ICLD- und ICTD-Parameter zwischen einem Referenzkanal und jedem anderen Kanal zu bestimmen. Dies ist in Fig. 4A dargestellt.

**[0058]** ICC-Parameter können auf verschiedene Arten und Weisen definiert werden. Allgemein gesagt kann man ICC-Parameter in dem Codierer zwischen allen möglichen Kanalpaaren bestimmen, wie es in Fig. 4B dargestellt ist. Es wurde jedoch vorgeschlagen, nur ICC-Parameter zwischen den stärksten zwei Kanälen zu einem Zeitpunkt zu berechnen, wie es in Fig. 4C dargestellt ist, wo ein Beispiel gezeigt ist, bei dem zu einem Zeitpunkt ein ICC-Parameter zwischen den Kanälen 1 und 2 berechnet wird, und zu einem anderen Zeitpunkt ein ICC-Parameter zwischen den Kanälen 1 und 5 berechnet wird. Der Decodierer synthetisiert dann die Interkanalkorrelation zwischen den stärksten Kanälen in dem De-

coder und verwendet bestimmte heuristische Regeln zum Berechnen und Synthetisieren der Interkanalkohärenz für die restlichen Kanalpaare.

**[0059]** Bezuglich der Berechnung beispielsweise der Multiplikationsparameter  $a_1, a_N$  basierend auf den übertragenen ICLD-Parametern wird auf das AES-Convention-Paper Nr. 5574 Bezug genommen. Die ICLD-Parameter stellen eine Energieverteilung eines ursprünglichen Mehrkanalsignals dar. Ohne Verlust der Allgemeinheit wird es bevorzugt, wie es in Fig. 4A gezeigt, vier ICLD-Parameter zu nehmen, die die Energiedifferenz zwischen den jeweiligen Kanälen und dem vorderen linken Kanal darstellen. In dem Seiteninformationsverarbeitungsblock 122 werden die Multiplikationsparameter  $a_1, \dots, a_N$  von den ICLD-Parametern so abgeleitet, dass die gesamte Energie aller rekonstruierter Ausgangskanäle dieselbe ist (oder proportional zu der Energie des übertragenen Summensignals ist).

**[0060]** Bei dem in Fig. 7 gezeigten Ausführungsbeispiel wird auf die Frequenz-Zeit-Umsetzung, die durch die inversen Filterbanken IFB 129 von Fig. 6 erreicht werden, verzichtet. Es werden stattdessen die Spektraldarstellungen der einzelnen Kanäle am Eingang dieser inversen Filterbanken verwendet und der Kopfhörer-Signalverarbeitungs-Vorrichtung von Fig. 7 zugeführt, um ohne eine zusätzliche Frequenz/Zeit-Transformation die Bewertung der einzelnen Multi-Kanäle mit den jeweils zwei Filtern pro Multi-Kanal durchzuführen.

**[0061]** Im Hinblick auf eine komplette im Frequenzbereich stattfindende Verarbeitung sei darauf hingewiesen, dass dann der Multikanal-Decodierer, also z. B. die Filterbank 125 von Fig. 6 und der Stereo-Codierer dieselbe Zeit/Frequenzauflösung haben sollen. Ferner wird es bevorzugt, ein und dieselbe Filterbank zu verwenden, was insbesondere auch dahingehend vorteilhaft ist, dass für die gesamte Verarbeitung, wie sie in Fig. 1 darstellt ist, nur eine einzige Filterbank benötigt wird. In diesem Fall ergibt sich eine besonders effiziente Verarbeitung, da die Transformation im Multikanal-Decodierer und im Stereo-Encodierer nicht berechnet werden müssen.

**[0062]** Die Eingangsdaten bzw. Ausgangsdaten beim erfindungsgemäßen Konzept sind daher vorzugsweise im Frequenzbereich codiert mittels Transformation/FILTERbank und sind nach psychoakustischen Vorgaben unter Ausnutzung von Verdeckungseffekten codiert, wobei insbesondere im Decodierer eine spektrale Darstellung der Signale vorliegen sollte. Beispiele hierfür sind MP3-Dateien, AAC-Dateien oder AC3-Dateien. Die Eingangsdaten bzw. Ausgangsdaten können jedoch auch durch Summen- und Differenzbildung codiert sein, wie es bei so genannten matrzierten Verfahren der Fall ist. Beispiele hierfür sind Dolby ProLogic, Logic7 oder Circle Surround. Die Daten insbesondere der Multikanaldarstellung können zusätzlich mit parametrischen Verfahren codiert sein, wie es bei MP3 Surround der Fall ist, wobei dieses Verfahren auf der BCC-Technik basiert.

**[0063]** Abhängig von den Gegebenheiten kann das erfindungsgemäße Verfahren zum Erzeugen in Hardware

oder in Software implementiert werden. Die Implementierung kann auf einem digitalen Speichermedium, insbesondere einer Diskette oder CD mit elektronisch auslesbaren Steuersignalen erfolgen, die so mit einem programmierbaren Computersystem zusammenwirken können, dass das Verfahren ausgeführt wird. Allgemein besteht die Erfindung somit auch in einem Computer-Programm-Produkt mit einem auf einem maschinenlesbaren Träger gespeicherten Programmcode zur Durchführung eines erfindungsgemäßen Verfahrens, wenn das Computer-Programm-Produkt auf einem Rechner abläuft. In anderen Worten ausgedrückt kann die Erfindung somit als ein Computer-Programm mit einem Programmcode zur Durchführung des Verfahrens realisiert werden, wenn das Computer-Programm auf einem Computer abläuft.

## Patentansprüche

1. Vorrichtung zum Erzeugen eines codierten Stereo-Signals eines Audiostücks oder Audiodatenstroms mit einem ersten Stereo-Kanal und einem zweiten Stereo-Kanal aus einer Multikanal-Darstellung des Audiostücks oder Audiodatenstroms, die Informationen über mehr als zwei Multi-Kanäle aufweist, mit folgenden Merkmalen:

einer Einrichtung (11) zum Bereitstellen der mehr als zwei Multi-Kanäle aus der Multikanal-Darstellung;

wobei die Multikanal-Darstellung einen oder mehrere Basiskanäle sowie Parameterinformationen zum Berechnen der Multi-Kanäle aus einem oder mehreren Basiskanälen aufweist,

wobei die Einrichtung (11) zum Bereitstellen ausgebildet ist, um aus dem einen oder den mehreren Basiskanälen und den Parameterinformationen die wenigstens drei Multi-Kanäle zu berechnen.

einer Einrichtung (12) zum Durchführen einer Kopfhörer-Signalverarbeitung, um ein uncodiertes Stereo-Signal mit einem uncodierten ersten Stereo-Kanal (10a) und einem uncodierten zweiten Stereo-Kanal (10b) zu erzeugen, wobei die Einrichtung (12) zum Durchführen ausgebildet ist,

um jeden Multi-Kanal mit einer ersten Filterfunktion ( $H_{iL}$ ), die von einer virtuellen Position eines Lautsprechers zum Wiedergeben des Multi-Kanals und einer virtuellen ersten Ohrposition eines Hörers abgeleitet ist, für den ersten Stereo-Kanal und mit einer zweiten Filterfunktion ( $H_{iR}$ ), die von einer virtuellen Position des Lautsprechers und einer virtuellen zweiten Ohrposition des Hörers abgeleitet ist, für den zweiten Stereo-Kanal zu bewerten, um für jeden Multi-Kanal einen ersten bewerteten Kanal und einen zweiten bewerteten Kanal zu erzeugen, wobei die beiden virtuellen Ohrpositionen des Hörers un-

terschiedlich sind,  
um die bewerteten ersten Kanäle aufzuaddieren (22), um den uncodierten ersten Stereo-Kanal (10a) zu erhalten, und  
um die bewerteten zweiten Kanäle aufzuaddieren (23), um den uncodierten zweiten Stereo-Kanal (10b) zu erhalten; und  
einem Stereo-Codierer (13) zum Codieren des ersten uncodierten Stereo-Kanals (10a) und des uncodierten zweiten Stereo-Kanals (10b), um das codierte Stereo-Signal (14) zu erhalten, wobei der Stereo-Codierer derart ausgebildet ist, dass eine Datenrate, die zum Übertragen des codierten Stereo-Signals nötig ist, kleiner als eine Datenrate ist, die zum Übertragen des uncodierten Stereo-Signals nötig ist.

2. Vorrichtung nach Anspruch 1, bei der die Einrichtung (12) zum Durchführen ausgebildet ist, um die erste Filterfunktion ( $H_{iL}$ ), die Direktschall, Reflexionen und diffusen Nachhall berücksichtigt, und die zweite Filterfunktion ( $H_{iR}$ ), die Direktschall, Reflexionen und diffusen Nachhall berücksichtigt, zu verwenden.

3. Vorrichtung nach Anspruch 2, bei der die erste und die zweite Filterfunktion einer Filterimpulsantwort entsprechen, die einen Peak bei einem kleinen Zeitwert, der den Direktschall darstellt, mehrere kleinere Peaks bei mittleren Zeitwerten, die die Reflexionen darstellen, und einen kontinuierlichen Bereich, der nicht mehr nach einzelnen Peaks aufgelöst ist und den diffusen Nachhall darstellt, aufweist.

4. Vorrichtung nach Anspruch 1,  
bei der die Einrichtung (11) zum Bereitstellen ausgebildet ist, um ausgangsseitig eine blockweise Frequenzbereichsdarstellung für jeden Multi-Kanal zu liefern, und  
bei der die Einrichtung (12) zum Durchführen ausgebildet ist, um die blockweise Frequenzbereichsdarstellung mit einer Frequenzbereichs-Darstellung der ersten und der zweiten Filterfunktion zu bewerten.

5. Vorrichtung nach einem der vorhergehenden Ansprüche,  
bei der die Einrichtung (12) zum Durchführen ausgebildet ist, um eine blockweise Frequenzbereichsdarstellung des uncodierten ersten Stereo-Kanals und des uncodierten zweiten Stereo-Kanals zu liefern, und  
bei der der Stereo-Codierer (13) ein Transformationsbasiertes Codierer ist und ferner ausgebildet ist, um die blockweise Frequenzbereichsdarstellung des uncodierten ersten Stereo-Kanals und des uncodierten zweiten Stereo-Kanals ohne eine Umsetzung von der Frequenzbereichs-Darstellung in eine zeitliche Darstellung zu verarbeiten.

6. Vorrichtung nach einem der vorhergehenden Ansprüche,  
bei der der Stereo-Codierer (13) ausgebildet ist, um eine gemeinsame Stereo-Codierung (15) des ersten und des zweiten Stereokanals durchzuführen. 5
7. Vorrichtung nach einem der vorhergehenden Ansprüche,  
bei der der Stereo-Codierer (13) ausgebildet ist, um einen Block von Spektralwerten unter Verwendung einer psychoakustischen Maskierungsschwelle zu quantisieren (16) und einer Entropie-Codierung (17) zu unterziehen, um das codierte Stereo-Signal zu erhalten. 10
8. Vorrichtung nach einem der vorhergehenden Ansprüche,  
bei der die Einrichtung (11) zum Bereitstellen als BCC-Decodierer ausgebildet ist. 15
9. Vorrichtung nach einem der vorhergehenden Ansprüche,  
bei der die Einrichtung (11) zum Bereitstellen als Multikanal-Decodierer ausgebildet ist, der eine Filterbank mit mehreren Ausgängen aufweist,  
bei der die Einrichtung (12) zum Durchführen ausgebildet ist, um Signale an den Filterbank-Ausgängen mit der ersten und der zweiten Filterfunktion zu bewerten, und  
bei der der Stereo-Codierer (13) ausgebildet ist, um den im Frequenzbereich vorliegenden uncodierten ersten Stereo-Kanal und den im Frequenzbereich vorliegenden uncodierten zweiten Stereo-Kanal zu quantisieren (16) und einer Entropie-Codierung (17) zu unterziehen, um das codierte Stereo-Signal zu erhalten. 20
10. Verfahren zum Erzeugen eines codierten Stereo-Signals eines Audiostücks oder Audiodatenstroms mit einem ersten Stereo-Kanal und einem zweiten Stereo-Kanal aus einer Multikanal-Darstellung des Audiostücks oder Audiodatenstroms, die Informationen über mehr als zwei Multi-Kanäle aufweist, mit folgenden Schritten: 25
- Bereitstellen (11) der mehr als zwei Multi-Kanäle aus der Multikanal-Darstellung; 30
- wobei die Multikanal-Darstellung einen oder mehrere Basiskanäle sowie Parameterinformationen zum Berechnen der Multi-Kanäle aus einem oder mehreren Basiskanälen aufweist,  
wobei beim Bereitstellen aus dem einen oder den mehreren Basiskanälen und den Parameterinformationen die wenigstens drei Multi-Kanäle berechnet werden; 35
- Durchführen (12) einer Kopfhörer-Signalverarbeitung, um ein uncodiertes Stereosignal mit einem un- 40
- 45
- 50
- 55

codierten ersten Stereo-Kanal (10a) und einem uncodierten zweiten Stereo-Kanal (10b) zu erzeugen, wobei der Schritt des Durchführens (12) folgende Schritte aufweist:

Bewerten jedes Multi-Kanals mit einer ersten Filterfunktion ( $H_{1L}$ ), die von einer virtuellen Position eines Lautsprechers zum Wiedergeben des Multi-Kanals und einer virtuellen ersten Ohrposition eines Hörers abgeleitet ist, für den ersten Stereo-Kanal und mit einer zweiten Filterfunktion ( $H_{1R}$ ), die von einer virtuellen Position des Lautsprechers und einer virtuellen zweiten Ohrposition des Hörers abgeleitet ist, für den zweiten Stereo-Kanal, um für jeden Multi-Kanal einen ersten bewerteten Kanal und einen zweiten bewerteten Kanal zu erzeugen, wobei die beiden virtuellen Ohrpositionen des Hörers unterschiedlich sind,  
Aufaddieren (22) der bewerteten ersten Kanäle, um den uncodierten ersten Stereo-Kanal (10a) zu erhalten, und  
Aufaddieren (23) der bewerteten zweiten Kanäle, um den uncodierten zweiten Stereo-Kanal (10b) zu erhalten; und  
Stereo-Codieren (13) des ersten uncodierten Stereo-Kanals (10a) und des uncodierten zweiten Stereo-Kanals (10b), um das codierte Stereo-Signal (14) zu erhalten, wobei der Schritt des Stereo-Codierens derart ausgeführt wird, dass eine Datenrate, die zum Übertragen des codierten Stereo-Signals nötig ist, kleiner als eine Datenrate ist, die zum Übertragen des uncodierten Stereo-Signals nötig ist.

11. Computer-Programm mit einem Programmcode zum Durchführen des Verfahrens zum Erzeugen eines codierten Stereo-Signals nach Patentanspruch 10, wenn das Computer-Programm auf einem Rechner abläuft.

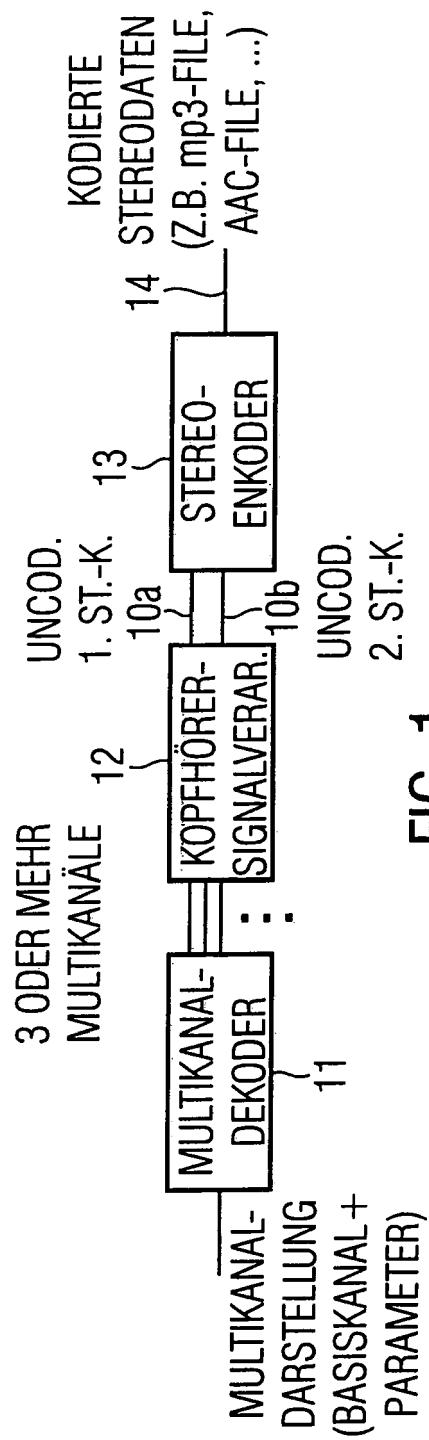


FIG. 1

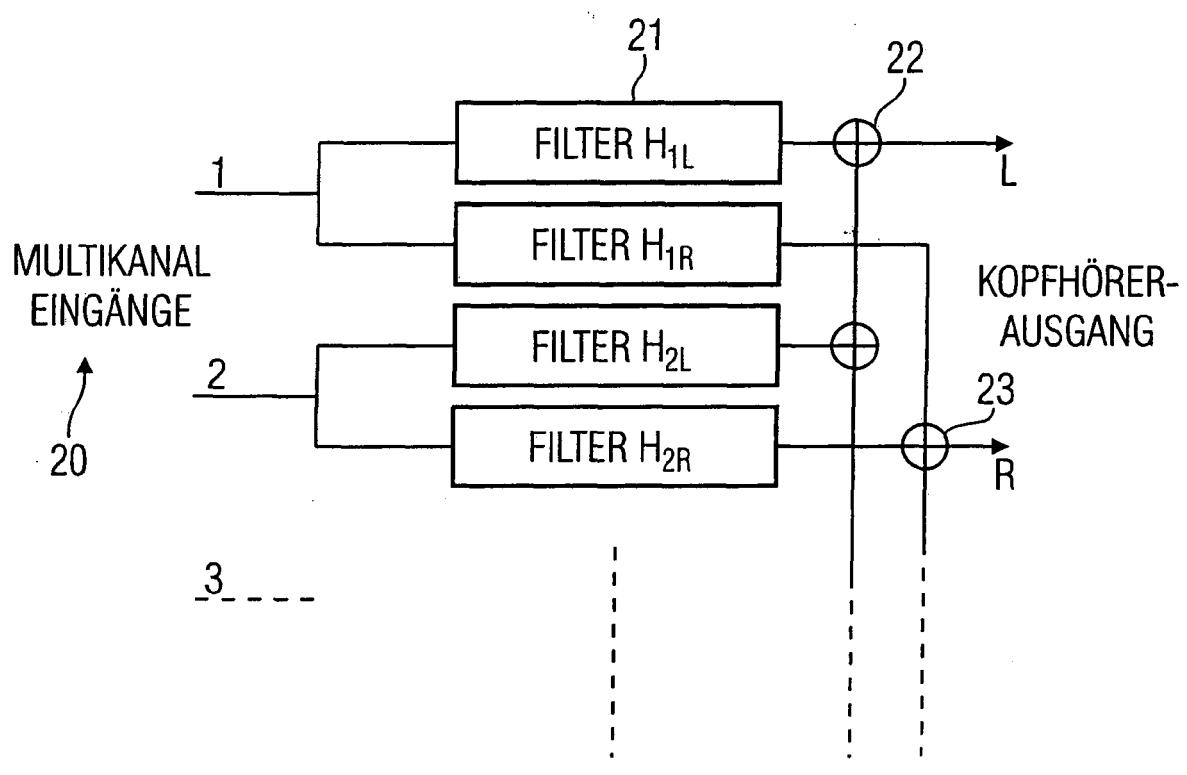


FIG. 2

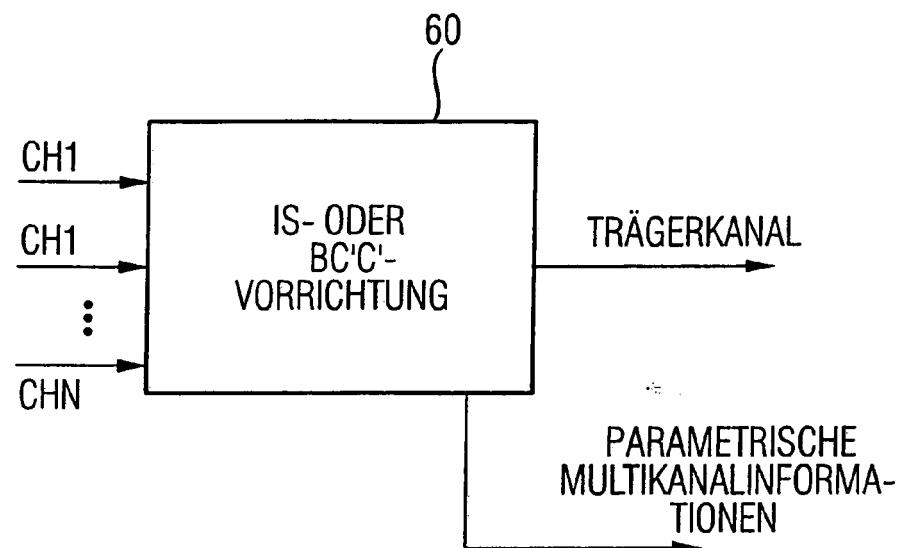


FIG. 3

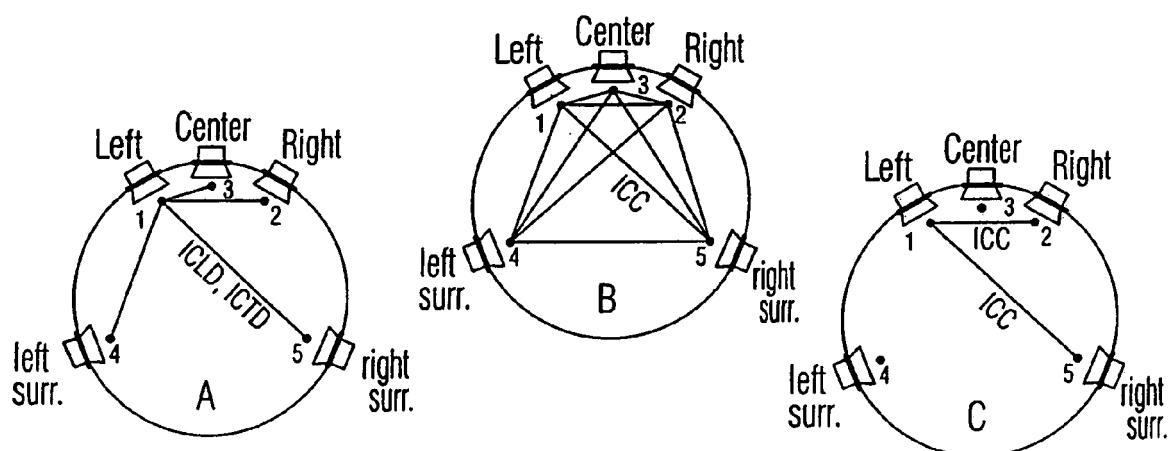


FIG. 4

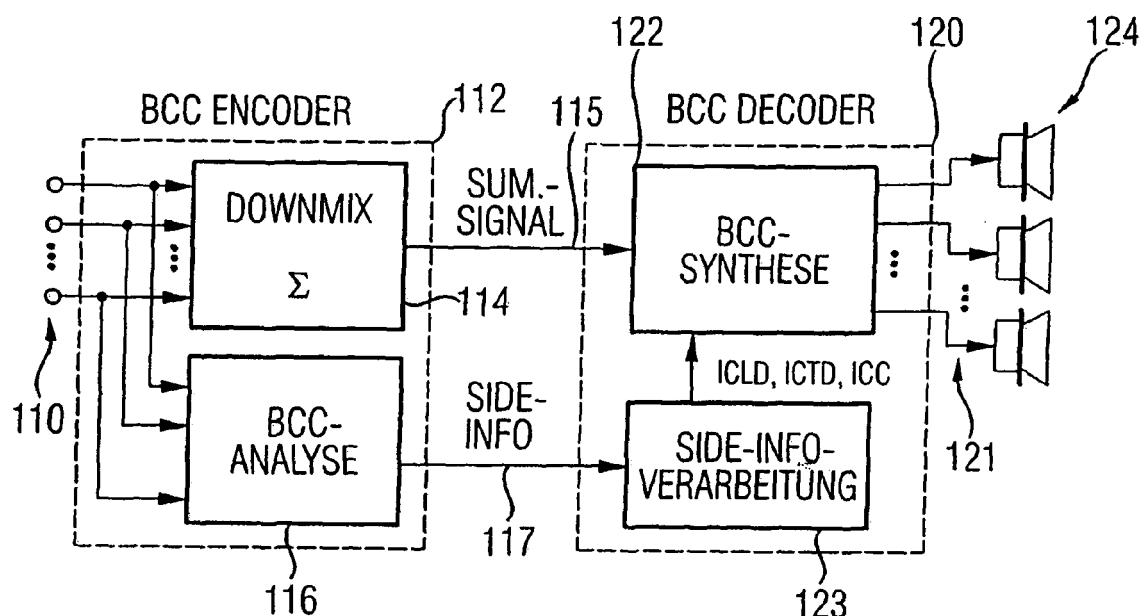


FIG. 5

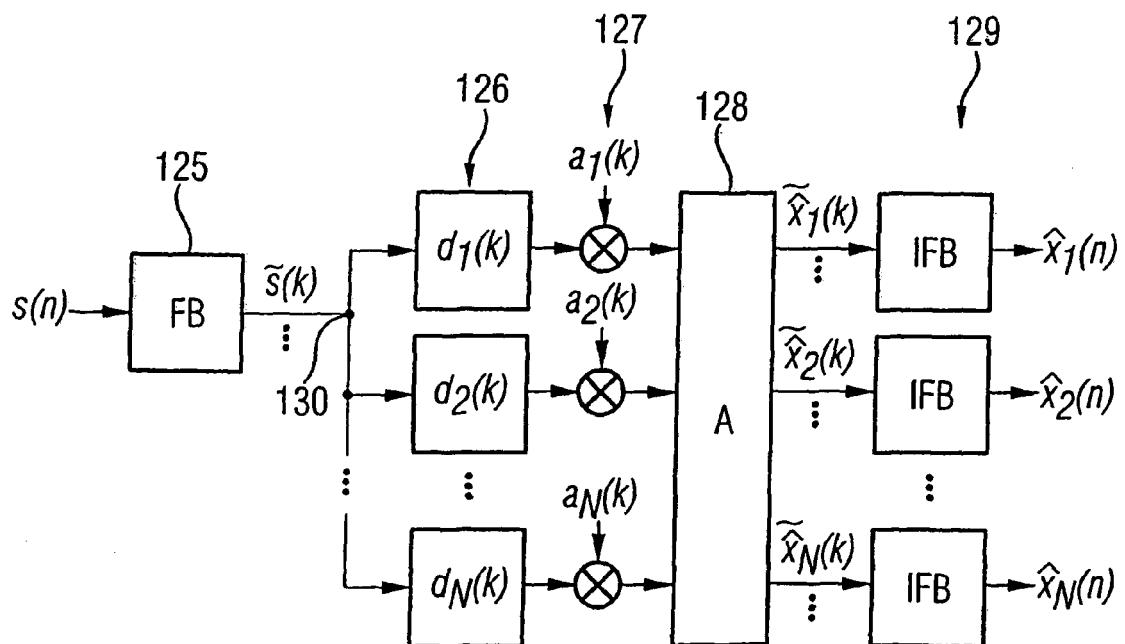


FIG. 6

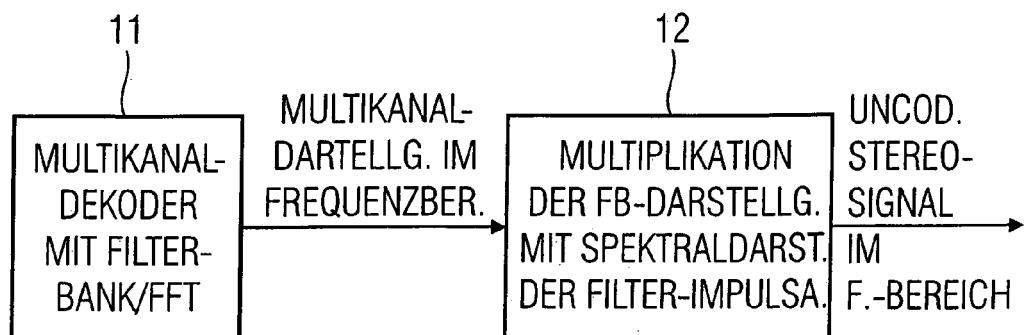


FIG. 7

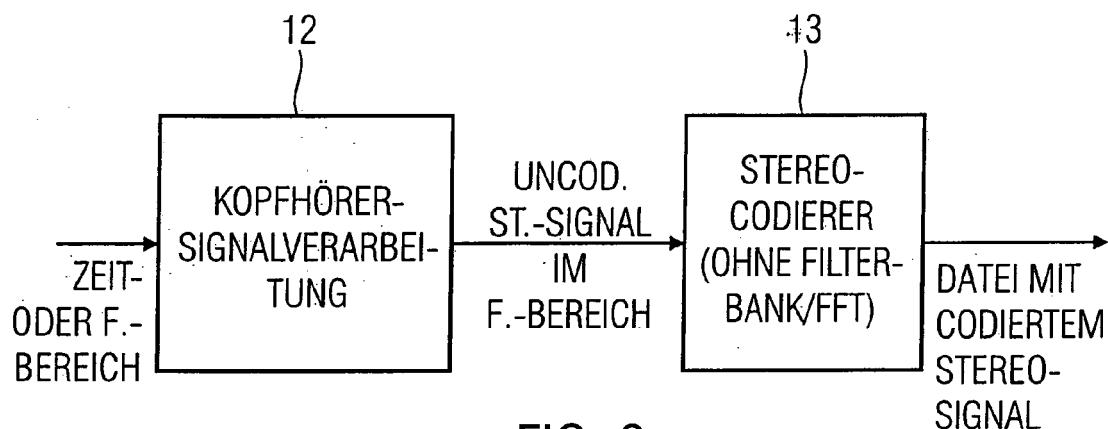


FIG. 8

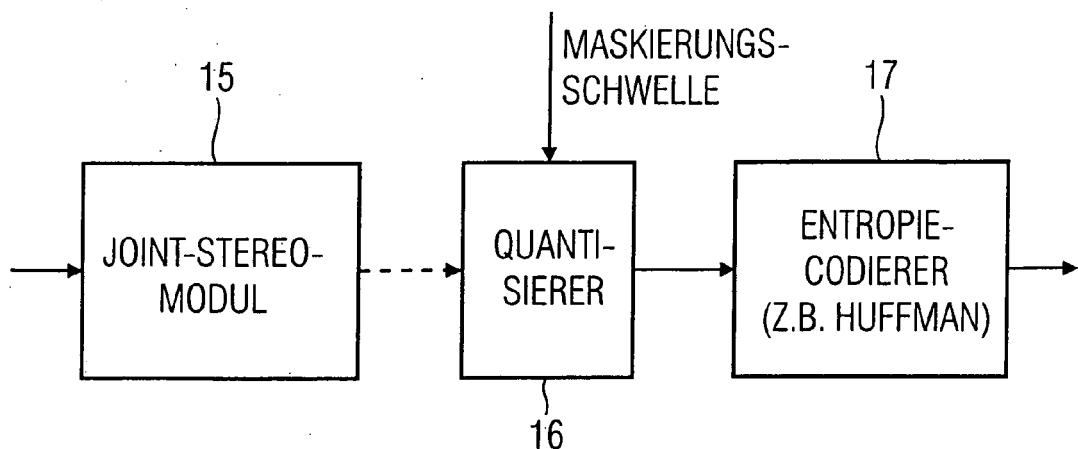


FIG. 9

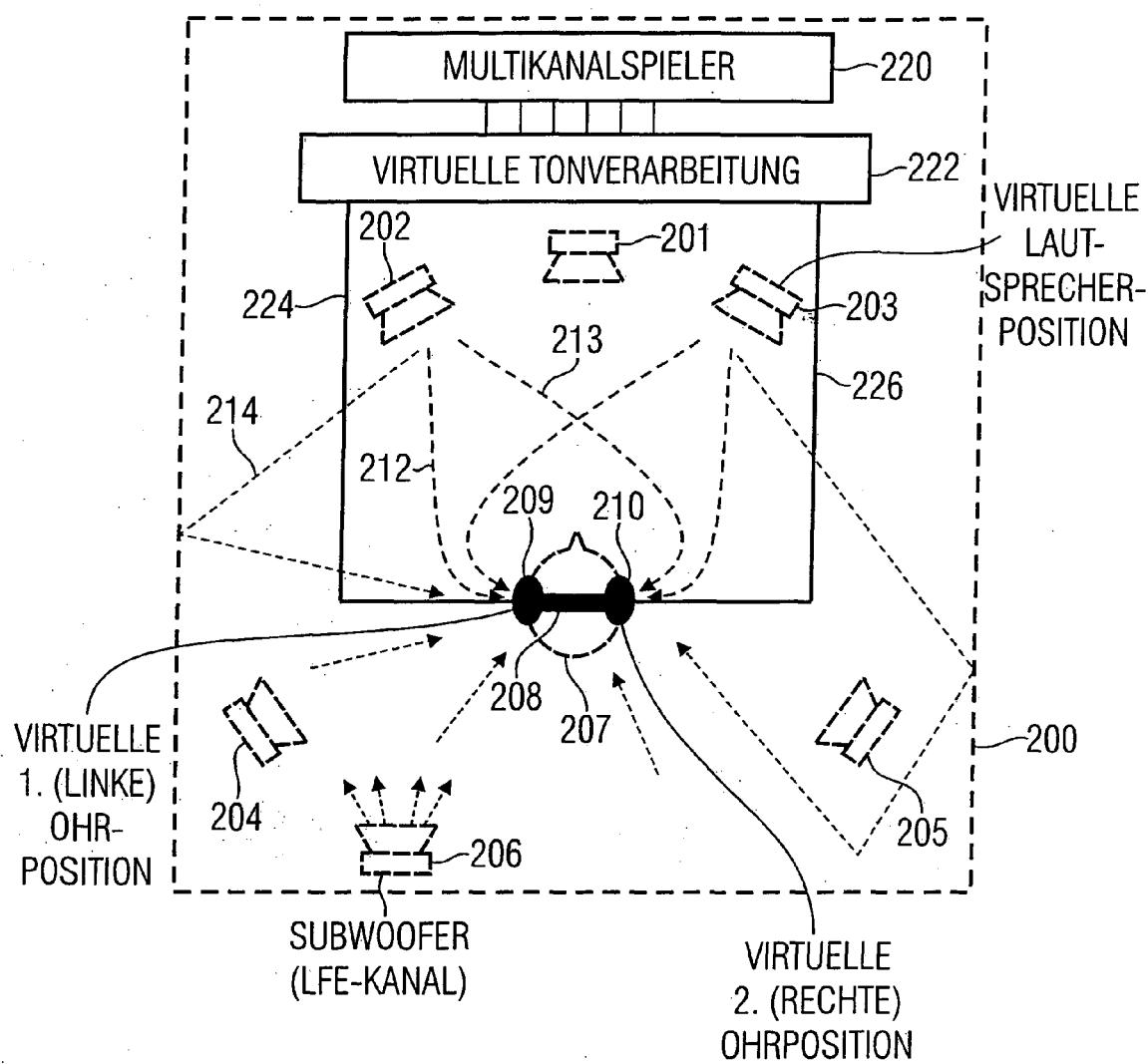
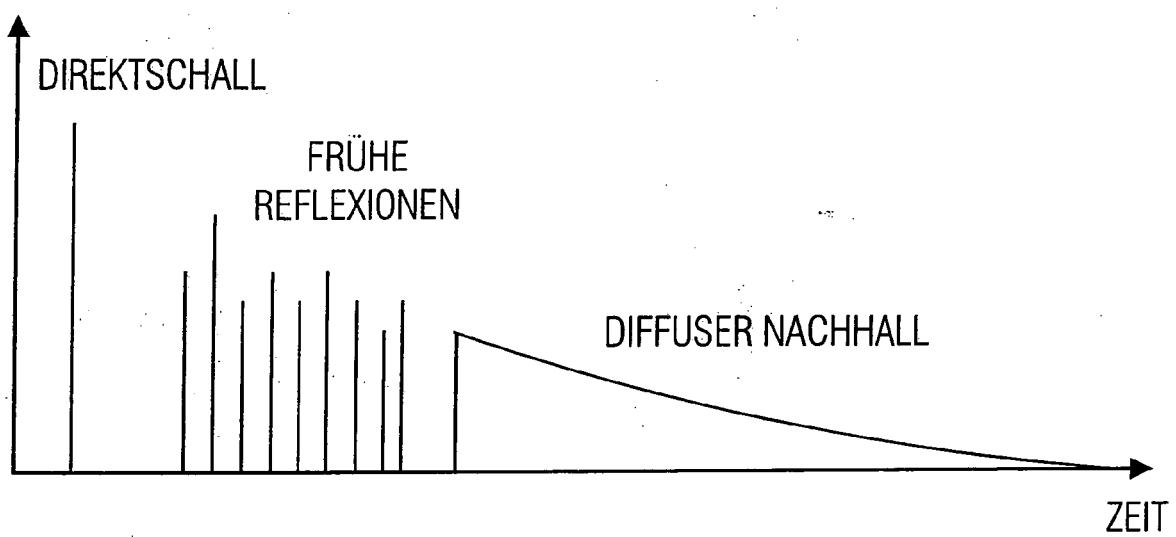


FIG. 10



**IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE**

*Diese Liste der vom Anmelder aufgeführten Dokumente wurde ausschließlich zur Information des Lesers aufgenommen und ist nicht Bestandteil des europäischen Patentdokumentes. Sie wurde mit größter Sorgfalt zusammengestellt; das EPA übernimmt jedoch keinerlei Haftung für etwaige Fehler oder Auslassungen.*

**In der Beschreibung aufgeführte Patentdokumente**

- WO 9949574 A [0002]
- WO 9914983 A [0002]
- WO 9401933 A [0022]
- US 20030219130 A1 [0045]
- US 20030026441 A1 [0045]
- US 20030035553 A1 [0045]

**In der Beschreibung aufgeführte Nicht-Patentliteratur**

- **J. Herre ; K.H. Brandenburg ; D. Lederer.** *Intensity Stereo Coding*, Februar 1994 [0038]
- **T. Faller ; F. Baumgarte.** *Binaural Cue Coding applied to stereo and multichannel audio compression*, Mai 2002 [0040]
- **T. Faller ; F. Baumgarte.** Binaural Cue Coding. Part II: Schemes and Applications. *IEEE Trans. On Audio and Speech Proc*, November 2003, vol. 11 (6 [0045]