



(11)

EP 3 275 212 B1

(12)

EUROPÄISCHE PATENTSCHRIFT

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
26.06.2019 Patentblatt 2019/26

(51) Int Cl.:
H04S 5/00 (2006.01) H04S 3/00 (2006.01)
H04S 3/02 (2006.01)

(21) Anmeldenummer: **16713352.9**

(86) Internationale Anmeldenummer:
PCT/EP2016/056163

(22) Anmeldetag: **21.03.2016**

(87) Internationale Veröffentlichungsnummer:
WO 2016/156091 (06.10.2016 Gazette 2016/40)

(54) VERFAHREN ZUR ANALYSE UND DEKOMPOSITION VON STEREOAUDIOSIGNALEN

METHOD FOR ANALYSING AND DECOMPOSING STEREO AUDIO SIGNALS

PROCÉDÉ D'ANALYSE ET DE DÉCOMPOSITION DE SIGNAUX AUDIO STÉRÉOPHONIQUES

(84) Benannte Vertragsstaaten:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(74) Vertreter: **Maiwald Patent- und Rechtsanwaltsgesellschaft mbH**
Elisenhof
Elisenstraße 3
80335 München (DE)

(30) Priorität: **27.03.2015 DE 102015104699**

(43) Veröffentlichungstag der Anmeldung:
31.01.2018 Patentblatt 2018/05

(56) Entgegenhaltungen:
WO-A1-2010/028784 DE-B4-102012 017 296
US-A1- 2011 116 638 US-A1- 2013 170 649

(73) Patentinhaber:
• **Helmut-Schmidt-Universität**
22043 Hamburg (DE)
• **Hamburg Innovation GmbH**
21079 Hamburg (DE)

- **LEE KANGEUN ET AL: "Virtual 5.1 Channel Reproduction of Stereo Sound for Mobile Devices", AES CONVENTION 132; APRIL 2012, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 26. April 2012 (2012-04-26), XP040574620,**
- **GOODWIN M M ET AL: "Primary-Ambient Signal Decomposition and Vector-Based Localization for Spatial Audio Coding and Enhancement", 2007 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING 15-20 APRIL 2007 HONOLULU, HI, USA, IEEE, PISCATAWAY, NJ, USA, 15. April 2007 (2007-04-15), Seiten I-9, XP031462785, ISBN: 978-1-4244-0727-9**

(72) Erfinder:
• **KRAFT, Sebastian**
22083 Hamburg (DE)
• **FINK, Marco**
22307 Berlin (DE)
• **MIETH, Martin**
04155 Leipzig (DE)
• **ZÖLZER, Udo**
21629 Neu Wulmstorf (DE)

EP 3 275 212 B1

Anmerkung: Innerhalb von neun Monaten nach Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents im Europäischen Patentblatt kann jedermann nach Maßgabe der Ausführungsordnung beim Europäischen Patentamt gegen dieses Patent Einspruch einlegen. Der Einspruch gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist. (Art. 99(1) Europäisches Patentübereinkommen).

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Analyse und Dekomposition eines Stereo-Audiosignals sowie ein Verfahren zum Erzeugen eines Mehrkanaludiosignals.

Stand der Technik

[0002] Mit der Aufzeichnung eines Stereoaudiosignals, wobei in der Regel ein erstes Audiosignal für eine linke Wiedergabeeinrichtung und ein zweites Audiosignal für eine rechte Wiedergabeeinrichtung verwendet wird, lässt sich der Eindruck erzeugen, dass Phantomschallquellen auf einen Hörbereich zwischen dem Hörer und den beiden Wiedergabeeinrichtungen verteilt sind.

[0003] Dabei liefert der Pegelunterschied zwischen dem ersten und dem zweiten Audiosignal primär die Information, aus welcher azimutalen Richtung relativ zum Hörer der Schall zu kommen scheint. Diese Information ist lediglich ein-dimensional und kann daher naturgemäß keine realistische Reproduktion von Räumlichkeit herstellen. Zudem ist der Azimutwinkel der möglichen Positionierung von Phantomschallquellen auf den Bereich beschränkt, der durch eine erste Verbindungsstrecke zwischen dem Hörer und der linken Wiedergabeeinrichtung und durch eine zweite Verbindungsstrecke zwischen dem Hörer und der rechten Wiedergabeeinrichtung aufgespannt wird. Des Weiteren ist es nur mit zwei Wiedergabeeinrichtungen nicht möglich Räumlichkeit zu simulieren, da hierfür der Schall aus allen Raumrichtungen abgestrahlt und auf den Hörer treffen müsste.

[0004] Mehrkanaludiosysteme mit beispielsweise fünf oder sieben Wiedergabeeinrichtungen vermitteln dem Hörer daher einen deutlich detaillierteren räumlichen Eindruck. Dieser Zusatznutzen liegt jedoch im Wesentlichen brach, wenn die Aufnahme nur als Stereoaudiosignal zur Verfügung steht.

[0005] Aus der DE 10 2012 017 296 B4 ist ein Verfahren zum Erzeugen eines Mehrkanaludiosignals aus einem Stereoaudiosignal bekannt. Damit lassen sich gerichtete Direktschallanteile und diffuse Umgebungsschallanteile in einem Stereoaudiosignal trennen sowie die Richtungsinformation der Direktschallanteile bestimmen, um anschließend alle Signalbestandteile auf einer Mehrkanalwiedergabeeinrichtung abzuspielen. Allerdings ist dieses Verfahren sehr rechenaufwändig.

[0006] Die WO 2010/028 784 A1 offenbart ein weiteres Verfahren zur Zerlegung eines Stereoaudiosignals in einen diffusen Umgebungsschallanteil einerseits und in gerichtete Direktschallanteile andererseits, wobei auch die Richtungen geschätzt werden, aus denen die Direktschallanteile zu kommen scheinen.

Aufgabe und Lösung

[0007] Aufgabe der vorliegenden Erfindung ist daher, bei gleichbleibender oder besserer Klangqualität die in einem Stereoaudiosignal enthaltene räumliche Information über die Anordnung der Schallquellen mit einem geringeren Rechenaufwand zu rekonstruieren.

[0008] Diese Aufgabe wird erfindungsgemäß gelöst durch Verfahren zur Analyse gemäß Haupt- und Nebenanspruch sowie durch ein Verfahren zum Erzeugen eines Mehrkanaludiosignals gemäß weiterem Nebenanspruch. Weitere vorteilhafte Ausgestaltungen ergeben sich aus den daraus rückbezogenen Unteransprüchen.

Gegenstand der Erfindung

[0009] Im Rahmen der Erfindung wurde ein Verfahren zur Analyse und Dekomposition eines Stereoaudiosignals entwickelt. Dieses Stereoaudiosignal weist ein erstes Audiosignal für eine linke Wiedergabeeinrichtung und ein zweites Audiosignal für eine rechte Wiedergabeeinrichtung auf.

[0010] Erfindungsgemäß sieht das Verfahren folgende Schritte vor:

Zunächst wird das erste Audiosignal in eine erste Zeit-Frequenz-Darstellung überführt. Das zweite Audiosignal wird in eine zweite Zeit-Frequenz-Darstellung überführt. Die Überführung der Audiosignale in die Zeit-Frequenz-Darstellung kann mit beliebigen Verfahren erfolgen. Bevorzugt wird die Kurzzeit-Fourier-Transformation (STFT) verwendet.

[0011] Nun wird eine erste Gleichung aufgestellt, die die erste Zeit-Frequenz-Darstellung in Beziehung setzt zum Produkt eines ersten zeit- und frequenzabhängigen Panning-Koeffizienten mit dem zeit- und frequenzabhängigen Signal einer in einem Hörbereich zwischen der linken Wiedergabeeinrichtung und der rechten Wiedergabeeinrichtung angeordneten Direktschallquelle. Es wird eine zweite Gleichung aufgestellt, die die zweite Zeit-Frequenz-Darstellung in Beziehung setzt zum Produkt eines zweiten zeit- und frequenzabhängigen Panning-Koeffizienten mit dem gleichen Signal der gleichen Direktschallquelle. Dabei sind die Panning-Koeffizienten dazu ausgebildet, die Direktschallquelle in dem Hörbereich zu positionieren.

[0012] Es werden nun die Panning-Koeffizienten und/oder ein Positionskoeffizient, der der Differenz der Quadrate der Panning-Koeffizienten entspricht, als Lösungen des aus beiden Gleichungen gebildeten Gleichungssystems ermittelt.

Zu dem Stereoaudiosignal hat in der Regel eine Vielzahl unabhängiger Schallquellen beigetragen. Derjenige Anteil des ersten und des zweiten Audiosignals, der dem Richtungshören zugänglich ist, setzt sich also aus Beiträgen dieser einzelnen Schallquellen zusammen. Jeder dieser einzelnen Beiträge ist das Produkt einer zeit- und frequenzabhängigen komplexen Amplitude und eines Panning-Koeffizienten, der von der Positionierung der Schallquelle relativ zum Hörer abhängt. Das linke und das rechte Audiosignal sind, wenn man jeweils von Umgebungssignalen absieht, jeweils Summen über solche Einzelbeiträge. Da die Umgebungssignale diffus, über alle Raumrichtungen gleichverteilt und außerdem klein gegenüber dem Direktsignal sind, können sie in dem Gleichungssystem für die Ermittlung der Panning-Koeffizienten unberücksichtigt bleiben. Das Gleichungssystem wird dadurch deutlich einfacher lösbar.

[0013] Bei der Aufstellung des Gleichungssystems wird die vereinfachende Annahme gemacht, dass alle gleichzeitig aktiven Schallquellen zu einer einzigen Schallquelle mit zeit- und frequenzabhängiger komplexer Amplitude zusammengefasst werden können. Dies ist möglich, da bei einer ausreichend großen Zeit-Frequenz-Auflösung der Zeit-Frequenz-Darstellung davon auszugehen ist, dass zu einem bestimmten Zeitpunkt und in einem bestimmten Frequenzband nur eine einzige dominante Schallquelle existiert.

[0014] Dabei ist die komplexe Amplitude dieser zusammengefassten Schallquelle richtungsunabhängig. Die Richtungsabhängigkeit steckt allein in den Panning-Koeffizienten. Durch die Zusammenfassung der einzelnen Schallquellen lassen sich nun der erste und der zweite Panning-Koeffizient jeder Schallquelle zu einem Paar zeit- und frequenzabhängiger Panning-Koeffizienten für die zusammengefasste Schallquelle vereinen.

[0015] Unter der Annahme, dass der erste und der zweite Panning-Koeffizient miteinander verknüpft sind, lässt sich das Gleichungssystem mathematisch umformen, und die Panning-Koeffizienten lassen sich aus dem ersten und zweiten Kanal des Stereosignals ermitteln. Die Verknüpfung zwischen den beiden Panning-Koeffizienten ermöglicht es, das Gleichungssystem durch einfaches mathematisches Umformen zu lösen und für die Panning-Koeffizienten eine geschlossene Formel in den Zeit-Frequenz-Darstellungen des linken und des rechten Audiosignals anzugeben. Im laufenden Betrieb des Verfahrens sind Lösungen des Gleichungssystems also besonders schnell durch Einsetzen der Zeit-Frequenz-Darstellungen in die geschlossene Formel erhältlich.

[0016] Das Gleichungssystem wird unter der zusätzlichen Bedingung gelöst, dass die Summe der Quadrate der Panning-Koeffizienten konstant ist. Die Summe dieser Quadrate ist bei dem in der Musikproduktion üblicherweise verwendeten Constant-Power-Panning gleich 1. Diese führt dazu, dass die Schallquelle unabhängig von ihrer Position im Hörbereich gleich laut wahrgenommen wird.

[0017] Die Panning-Koeffizienten enthalten die vollständige Information, auf welcher Frequenz das Signal zu welcher Zeit von welchem Ort im Hörbereich zu kommen scheint.

[0018] Da sich die einzelnen Schallquellen inkohärent überlagern und die Aufnahme des Stereoaudiosignals ebenfalls inkohärent erfolgt, ändert eine unterschiedliche Positionierung der Schallquellen im Hörbereich nur die Amplitude des aufgenommenen Stereoaudiosignals, nicht jedoch dessen Phase. Daher sind auch die Zeit-Frequenz-Darstellungen des ersten und zweiten Audiosignals in Phase mit der zeit- und frequenzabhängigen komplexen Amplitude der Direkt-schallquelle. Damit kürzen sich die Phasenterme aus dem beschriebenen Gleichungssystem und nach umstellen ergibt sich der erste Panning-Koeffizient als Wurzel aus dem Verhältnis des Betragsquadrats der Zeit-Frequenz-Darstellung des ersten Audiosignals (Zähler) und der Summe der Betragsquadrate der Zeit-Frequenz-Darstellung des ersten und zweiten Audiosignals (Nenner). Analog ergibt sich der zweite Panning-Koeffizient als Wurzel aus dem Verhältnis des Betragsquadrats der Zeit-Frequenz-Darstellung des zweiten Audiosignals (Zähler) und der Summe der Betragsquadrate der Zeit-Frequenz-Darstellung des ersten und zweiten Audiosignals (Nenner).

[0019] Der Positionskoeffizient wird aus dem Verhältnis der Differenz der Betragsquadrate beider Zeit-Frequenz-Darstellungen zur Summe der Betragsquadrate beider Zeit-Frequenz-Darstellungen ermittelt.

[0020] Eine alternative Ausgestaltung der Erfindung startet ebenfalls von einem ersten Audiosignal für eine linke Wiedergabeeinrichtung und einem zweiten Audiosignal für eine rechte Wiedergabeeinrichtung. Das erste Audiosignal wird in eine erste Zeit-Frequenz-Darstellung überführt, und das zweite Audiosignal wird in eine zweite Zeit-frequenz-Darstellung überführt.

[0021] In dieser Ausgestaltung wird aus der ersten Zeit-Frequenz-Darstellung die zeit- und frequenzabhängige Leistung des ersten Audiosignals ermittelt, und aus der zweiten Zeit-Frequenz-Darstellung wird die zeit- und frequenzabhängige Leistung des zweiten Audiosignals ermittelt. Dementsprechend werden auch die Gleichungen für die Panning-Koeffizienten abgeändert.

[0022] Es wird eine erste Gleichung aufgestellt, die die zeit- und frequenzabhängige Leistung des ersten Audiosignals in Beziehung setzt zum Produkt des Quadrats eines ersten zeit- und frequenzabhängigen Panning-Koeffizienten mit der zeit- und frequenzabhängigen Leistung einer in einem Hörbereich zwischen der linken Wiedergabeeinrichtung und der rechten Wiedergabeeinrichtung angeordneten Direktschallquelle.

[0023] Es wird eine zweite Gleichung aufgestellt, die die zeit- und frequenzabhängige Leistung des zweiten Audiosignals in Beziehung setzt zum Produkt des Quadrats eines zweiten zeit- und frequenzabhängigen Panning-Koeffizienten mit der gleichen zeit- und frequenzabhängigen Leistung der gleichen Direktschallquelle.

[0024] Analog zu dem oben beschriebenen ersten Ansatz für das Gleichungssystem, in dem die Gleichungen die Zeit-

Frequenz-Darstellungen mit dem Signal der Direktschallquelle verknüpfen, sind die Panning-Koeffizienten dazu ausgebildet, die Direktschallquelle in dem Hörbereich zu positionieren. Die Panning-Koeffizienten, und/oder ein Positionskoeffizient, der dem Verhältnis einer Differenz der Panning-Koeffizienten zur Summe der Panning-Koeffizienten entspricht, werden als Lösungen des aus beiden Gleichungen gebildeten Gleichungssystems ermittelt.

[0025] Die Aufstellung des Gleichungssystems in Leistungen, statt direkt in Zeit-Frequenz-Darstellungen und Signal der Direktschallquelle, ist dadurch motiviert, dass das Panning ein reines Amplitudenpanning ist. Folglich sind beide Audiosignale in Phase mit dem Signal der Direktschallquelle. Wenn die Zeit-Frequenz-Darstellungen beispielsweise mit einer Kurzzeit-Fourier-Transformation (STFT) gewonnen wurden, lässt sich eine Leistung unmittelbar als Betragsquadrat des zugehörigen Leistungsdichtespektrums ausdrücken. Der Ansatz über die Leistungen ist dann äquivalent zum Ansatz über die Zeit-Frequenz-Darstellungen und das Signal der Direktschallquelle.

[0026] Der Ansatz über die Leistungen bietet jedoch den zusätzlichen Vorteil, dass er allgemeiner ist. Er ist auch dann anwendbar, wenn keine 1:1-Transformation der zeitabhängigen Audiosignale in einen Frequenzbereich vorliegt, sondern lediglich eine Aufspaltung dieser Audiosignale in mehrere zeitabhängige Signale, die zu den Beiträgen bestimmter Frequenzbänder korrespondieren. Eine solche Aufspaltung kann beispielsweise mit einer Filterbank hergestellt werden. Eine Filterbank enthält typischerweise mehrere parallel geschaltete Bandpassfilter, die jeweils den Anteil des Signals, der in einen bestimmten Frequenzband fällt, passieren lassen. Das Signal am Ausgang eines jeden dieser Bandpassfilter ist ein zeitabhängiges Signal. Die Gesamtheit aller dieser Signale bildet zusammen mit der Information, zu welchem Frequenzband ein jedes Signal korrespondiert, eine Zeit-Frequenz-Darstellung.

[0027] Eine derartige Zeit-Frequenz-Darstellung kann auf diese Weise zum Einen schneller und einfacher erhalten werden als mit der Kurzzeit-Fourier-Transformation (STFT). Beispielsweise können Bandpassfilter niedriger Ordnung und mit geringer Gruppenlaufzeit verwendet werden. Zum Anderen erleichtert eine derartige Zeit-Frequenz-Darstellung auch die frequenzabhängige Bearbeitung des Signals. Beispielsweise lässt sich die Frequenzauflösung variieren, indem ein weniger interessierender Frequenzbereich mit einem breiten Bandpassfilter abgedeckt wird, während ein besonders interessierender Frequenzbereich mit vielen schmalen Bandpassfiltern abgedeckt wird. Demgegenüber ist bei der Kurzzeit-Fourier-Transformation die Frequenzauflösung immer ein äquidistantes Raster.

[0028] Es ist nicht erforderlich, dass eine geschlossene Formel für die Berechnung der jeweiligen zeit- und frequenzabhängigen Leistung aus den Zeit-Frequenz-Darstellungen der beiden Audiosignale existiert. Es ist beispielsweise auch möglich, diese Leistung auf numerischem Wege näherungsweise zu ermitteln. Beispielsweise kann die zeit- und frequenzabhängige Leistung mindestens eines Audiosignals zu einem interessierenden Zeitpunkt als gewichtete Summe aus der zeit- und frequenzabhängigen Leistung des Audiosignals zu einem früheren Zeitpunkt und dem Quadrat der Zeit-Frequenz-Darstellung dieses Audiosignals zu dem interessierenden Zeitpunkt ermittelt werden. Ist die Zeit in der Zeit-Frequenz-Darstellung beispielsweise diskretisiert, so kann der frühere Zeitpunkt insbesondere eine diskrete Zeiteinheit vor dem interessierenden Zeitpunkt liegen. Die momentane Leistung eines Audiosignals kann also beispielsweise über eine rekursive Mittelung aus der Zeit-Frequenz-Darstellung ermittelt werden.

[0029] Das Gleichungssystem wird unter der zusätzlichen Bedingung gelöst, dass die Summe der Quadrate der Panning-Koeffizienten konstant ist, wobei der Positionskoeffizient aus dem Verhältnis der Differenz der Wurzeln der zeit- und frequenzabhängigen Leistungen beider Audiosignale zur Summe der Wurzeln der zeit- und frequenzabhängigen Leistungen beider Audiosignale ermittelt wird.

[0030] Das Auflösen des Gleichungssystems zu den Panning-Koeffizienten erfolgt in völlig analoger Weise zum Ansatz über die Zeit-Frequenz-Darstellungen und das Signal der Direktschallquelle. Die Panning-Koeffizienten und ggfs. der Positionskoeffizient sind lediglich durch andere Größen ausgedrückt.

[0031] Vorteilhaft wird daher der erste Panning-Koeffizient als Wurzel aus dem Verhältnis der zeit- und frequenzabhängigen Leistung des ersten Audiosignals zur Summe der zeit- und frequenzabhängigen Leistungen beider Audiosignale ermittelt. Der zweite Panning-Koeffizient wird dementsprechend als Wurzel aus dem Verhältnis der zeit- und frequenzabhängigen Leistung des zweiten Audiosignals zur Summe der zeit- und frequenzabhängigen Leistungen beider Audiosignale ermittelt.

[0032] Vorteilhaft wird die zeit- und frequenzabhängige Leistung mindestens eines Audiosignals zu einem interessierenden Zeitpunkt als gewichtete Summe aus der zeit- und frequenzabhängigen Leistung des Audiosignals zu einem früheren Zeitpunkt und dem Quadrat der Zeit-Frequenz-Darstellung dieses Audiosignals zu dem interessierenden Zeitpunkt ermittelt.

[0033] Im Allgemeinen wird das Stereoaudiosignal nicht nur einen richtungsabhängigen Direktsignalanteil enthalten. Stattdessen werden das erste und das zweite Audiosignal jeweils mit einem diffusen Umgebungssignal überlagert sein. Daher werden in einer weiteren besonders vorteilhaften Ausgestaltung der Erfindung aus den Panning-Koeffizienten das Signal der Direktschallquelle (Direktsignal) und/oder zwei nicht richtungsabhängige, d.h. nicht mit der Direktschallquelle korrelierte, Umgebungssignale ermittelt. Dabei ist das erste Umgebungssignal nur in der Zeit-Frequenz-Darstellung des ersten Audiosignals enthalten, und das zweite Umgebungssignal ist nur in der Zeit-Frequenz-Darstellung des zweiten Audiosignals enthalten. Das Hörerlebnis wird genauer reproduziert, wenn nur das Direktsignal mit Hilfe der Panning-Koeffizienten in gerichteter Form wiedergegeben wird. Das diffuse Umgebungssignal sollte auch diffus wie-

dergegeben werden.

[0034] Vorteilhaft werden das Direktsignal und die Umgebungssignale mit einem iterativen Verfahren bestimmt ausgehend von einer Iterationsvorschrift, die das Direktsignal einer jeder Iteration, und/oder einen Beitrag zu diesem Signal, in Beziehung setzt zu den Umgebungssignalen der vorherigen Iteration. Beispielsweise kann in jeder Iteration die Lautstärke eines Beitrags zum Direktsignal als arithmetisches Mittel der Lautstärken beider Umgebungssignale der vorherigen Iteration festgelegt werden. Dies geht von der Annahme aus, dass das Direktsignal im ersten und zweiten Audiosignal mit gleicher Phase vorliegt und die Umgebungssignale dazu Phasenverschoben sind.

[0035] Die Näherung kann verfeinert werden, indem bei jeder Iteration die Panning-Koeffizienten aus den Umgebungssignalen der vorherigen Iterationen neu berechnet werden. Hierfür können beispielsweise die Umgebungssignale der vorherigen Iteration als Zeit-Frequenz-Darstellungen eines linken und eines rechten Audiosignals gewertet werden, so dass die Panning-Koeffizienten wie zuvor beschrieben durch Lösen eines Gleichungssystems berechnet werden können.

[0036] Vorteilhaft wird dann das erste Umgebungssignal bei jeder Iteration um einen Betrag korrigiert, der das Produkt aus dem neu berechneten ersten Panning-Koeffizienten mit dem Direktsignal, oder mit dem Signalbeitrag, gemäß der aktuellen Iteration ist. Analog wird das zweite Umgebungssignal bei jeder Iteration um einen Betrag korrigiert, der das Produkt aus dem neu berechneten zweiten Panning-Koeffizienten mit dem Direktsignal, oder mit dem Signalbeitrag, gemäß der aktuellen Iteration ist. Der Hintergedanke ist hierbei, dass die Lösung selbstkonsistent sein soll: Ein Signal, das sich im Nachhinein als mit dem Signal der Direktschallquelle korreliert und somit als Teil des Direktsignals erweist, kann offensichtlich nicht zum diffusen Umgebungssignal zählen.

[0037] Nach Durchlauf aller Iterationen ergibt sich das gesamte Direktsignal als Summe der in allen einzelnen Iterationen ermittelten Signalbeiträge. Da sowohl die iterativ berechneten Panning-Koeffizienten als auch das iterativ bestimmte Direktsignal nur jeweils Schätzungen sind, ist nicht garantiert, dass die Summe aus dem mit dem ersten Panning-Koeffizienten gewichteten Direktsignal und dem ersten Umgebungssignal genau den Wert der Zeit-Frequenz-Darstellung des ersten Audiosignals entspricht. Analog kann nicht garantiert werden, dass die Summe aus dem mit dem zweiten Panning-Koeffizienten gewichteten Direktsignal und dem zweiten Umgebungssignal genau den Wert der Zeit-Frequenz-Darstellung des zweiten Audiosignals reproduziert. Das Direktsignal und die Umgebungssignale gehorchen zusammen also nicht notwendigerweise dem Signalmodell, das der Aufteilung der Zeit-Frequenz-Darstellungen des ersten und des zweiten Audiosignals in jeweils einen gerichteten und einen diffusen Anteil zu Grunde lag. Daher ist es vorteilhaft, nicht unmittelbar die in der letzten Iteration ermittelten Umgebungssignale weiterzuverwenden, sondern das erste Umgebungssignal als Differenz aus der ersten Zeit-Frequenz-Darstellung und dem mit dem ersten Panning-Koeffizienten gemäß der ersten Iteration gewichteten Direktsignal zu ermitteln. Analog sollte das zweite Umgebungssignal als Differenz zwischen der zweiten Zeit-Frequenz-Darstellung und dem mit dem zweiten Panning-Koeffizienten gemäß der ersten Iteration gewichteten Direktsignal ermittelt werden.

[0038] Ein weiterer vorteilhafter Ansatz für die Ermittlung der nicht mit der Direktschallquelle korrelierten Umgebungssignale geht von der Annahme aus, dass beide Umgebungssignale zwar ähnlich klingen, aber durch unterschiedliche Ausbreitungswege und Reflexionen dekorreliert sind.

[0039] Es wird eine erste Gleichung aufgestellt, die die erste Zeit-Frequenz-Darstellung in Beziehung setzt zur Summe aus dem Produkt des ersten Panning-Koeffizienten mit dem zeit- und frequenzabhängigen Signal der Direktschallquelle sowie aus der Filterung eines einzigen gemeinsamen Umgebungssignals mit einer ersten Dekorrelationsfunktion.

[0040] Es wird eine zweite Gleichung aufgestellt, die die zweite Zeit-Frequenz-Darstellung in Beziehung setzt zur Summe aus dem Produkt des zweiten Panning-Koeffizienten mit dem zeit- und frequenzabhängigen Signal der Direktschallquelle sowie aus der Filterung des gemeinsamen Umgebungssignals mit einer zweiten Dekorrelationsfunktion.

[0041] Dabei kann die Filterung eines Signals mit einer Dekorrelationsfunktion beispielsweise durch eine Faltung des Signals mit der Dekorrelationsfunktion realisiert sein.

[0042] Das zeit- und frequenzabhängige Signal der Direktschallquelle, und/oder das gemeinsame Umgebungssignal, werden als Lösungen des aus den beiden Gleichungen gebildeten Gleichungssystems ermittelt.

[0043] Die Dekorrelationsfunktionen können mit verschiedenen nach Stand der Technik bekannten Verfahren initialisiert werden, um realistisch klingende dekorrelierte Signale zu erhalten. Typischerweise werden dazu die Funktionen so erzeugt, dass sich zufällige Frequenzgänge ergeben.

[0044] In einer Zeit-Frequenz-Darstellung lässt sich eine Filterung, und hier insbesondere eine Faltung, näherungsweise als frequenzbandweise Multiplikation mit der Dekorrelationsfunktion ausdrücken. Hierin kann die Dekorrelationsfunktion beispielsweise pro Frequenzband durch einen Verstärkungsfaktor und eine Phasendrehung repräsentiert werden.

[0045] Vorteilhaft wird somit das zeit- und frequenzabhängige Signal der Direktschallquelle als Differenz zwischen dem frequenzbandweisen Produkt der ersten Zeit-Frequenz-Darstellung mit der zweiten Dekorrelationsfunktion und dem frequenzbandweisen Produkt der zweiten Zeit-Frequenz-Darstellung mit der ersten Dekorrelationsfunktion, dividiert durch die Differenz zwischen dem frequenzbandweisen Produkt des ersten Panning-Koeffizienten mit der zweiten Dekorrelationsfunktion und dem frequenzbandweisen Produkt des zweiten Panning-Koeffizienten mit der ersten Dekorre-

lationsfunktion, ermittelt.

[0046] Vorteilhaft wird somit das gemeinsame Umgebungssignal als Differenz zwischen dem Produkt der zweiten Zeit-Frequenz-Darstellung mit dem ersten Panning-Koeffizienten und dem Produkt der ersten Zeit-Frequenz-Darstellung mit dem zweiten Panning-Koeffizienten, dividiert durch die Differenz zwischen dem frequenzbandweisen Produkt des ersten Panning-Koeffizienten mit der zweiten Dekorrelationsfunktion und dem frequenzbandweisen Produkt des zweiten Panning-Koeffizienten mit der ersten Dekorrelationsfunktion, ermittelt.

[0047] Im Rahmen der Erfindung wurde auch ein Verfahren zum Erzeugen eines Mehrkanalaudiosignals aus einem Stereoaudiosignal entwickelt. Dabei weist das Stereoaudiosignal ein erstes Audiosignal für eine linke Wiedergabeeinrichtung und ein zweites Audiosignal für eine rechte Wiedergabeeinrichtung auf.

[0048] Erfindungsgemäß wird das Stereoaudiosignal zunächst mit einem Verfahren gemäß der Erfindung analysiert. Anschließend wird aus den Panning-Koeffizienten eine Mehrzahl von Repanning-Koeffizienten ermittelt, wobei jeder dieser Repanning-Koeffizienten einen Tonkanal einer Mehrzahl von Tonkanälen des Mehrkanalaudiosignals zugeordnet wird. Dabei sind die Repanning-Koeffizienten für die Mehrzahl von Tonkanälen ausgeführt, eine Direktschallquelle in einem Hörbereich zwischen einer Mehrzahl von Wiedergabeeinrichtungen für das Mehrkanalaudiosignal zu positionieren. Das Signal der Direktschallquelle (Direktsignal) wird nun mit einem ersten Repanning-Koeffizienten verrechnet und einem ersten Tonkanal zugeordnet. Es wird mit einem zweiten Repanning-Koeffizienten verrechnet und einem zweiten Tonkanal zugeordnet. Es wird schließlich auch mit einem dritten Repanning-Koeffizienten verrechnet und einem dritten Tonkanal zugeordnet. Diese Signale dieser drei Tonkanäle können entweder direkt wiedergegeben oder für eine spätere Wiedergabe bzw. Weiterverarbeitung gespeichert werden.

[0049] Vorteilhaft wird das erste Umgebungssignal dem ersten Tonkanal additiv hinzugefügt, und das zweite Umgebungssignal wird dem dritten Tonkanal additiv hinzugefügt.

[0050] In einer weiteren vorteilhaften Ausgestaltung der Erfindung wird jeder Tonkanal in jeweils ein Wiedergabesignal des Mehrkanalaudiosignals überführt, wobei jedes Wiedergabesignal für jeweils eine Wiedergabeeinrichtung vorgesehen ist.

[0051] Die Ermittlung der Repanning-Koeffizienten stellt eine Umverteilung des richtungsabhängigen Direktsignals auf eine beliebige Lautsprecheranordnung dar. Das Umgebungssignal wird anschließend auf eine Auswahl von Lautsprechern additiv überlagert. Für das Repanning kann ein beliebiges Verfahren gemäß Stand der Technik verwendet werden, beispielsweise das Verfahren gemäß DE 10 2012 017 296 B4 oder auch das "vector base amplitude panning" gemäß (Ville Pulkki, "Virtual sound source positioning using vector based amplitude panning", Journal of the Audio Engineering Society, Vol. 45, Issue 6, pp. 456-466, June 1997).

[0052] In einer weiteren vorteilhaften Ausgestaltung der Erfindung lassen sich die extrahierten Direkt- und Umgebungsschallsignale nicht nur für die unmittelbare Wiedergabe des Stereo-Audiosignals als aufgewertetes Mehrkanalaudiosignal nutzen. Sie können beispielsweise für eine spätere Wiedergabe abgespeichert und/oder vor der Wiedergabe manipuliert werden, um das Hörerlebnis mit weiteren Effekten aufzuwerten.

[0053] Es wurde erkannt, dass bei der oben beschriebenen iterativen Berechnung des Direktsignals und der Umgebungssignale für eine gegen Unendlich strebende Iterationszahl beide Umgebungssignale gegen betragsmäßig gleiche Werte mit unterschiedlichen Vorzeichen streben. Sie sind also bis auf einen Phasenfaktor identisch. Mit dieser zusätzlichen Vereinfachung können das Direktsignal und die Umgebungssignale im laufenden Betrieb mit besonders wenig Rechenaufwand unmittelbar erhalten werden.

[0054] In einer weiteren besonders vorteilhaften Ausgestaltung der Erfindung wird somit das Signal der Direktschallquelle (Direktsignal) aus dem Verhältnis der Summe beider Zeit-Frequenz-Darstellungen der Audiosignale (Zähler) zur Summe beider Panning-Koeffizienten (Nenner) ermittelt wird. Weiterhin können auch die Umgebungssignale aus dem Verhältnis einer Differenz zwischen der Zeit-Frequenz-Darstellung des ersten Audiosignals, gewichtet mit dem zweiten Panning-Koeffizienten, und der Zeit-Frequenz-Darstellung des zweiten Audiosignals, gewichtet mit dem ersten Panning-Koeffizienten (Zähler), zur Summe beider Panning-Koeffizienten (Nenner) ermittelt werden.

Spezieller Beschreibungsteil

[0055] Nachfolgend wird der Gegenstand der Erfindung anhand von Figuren erläutert, ohne dass der Gegenstand der Erfindung hierdurch beschränkt wird. Es ist gezeigt:

Figur 1 Skizzenhafte Darstellung der vereinfachenden Annahme für die Ermittlung der Panning-Koeffizienten

Figur 2 Linearisierung der Azimut-Position durch Einführung des Positions-Koeffizienten Ψ

Figur 3 Repanning zwecks Wiedergabe als Mehrkanalaudiosignal

Figur 4 Zugang zu den Panning-Koeffizienten über Gleichungen in Leistungen

Figur 5 Bestimmung der Umgebungssignale und des Direktsignals aus den Panning-Koeffizienten über ein weiteres Gleichungssystem

[0056] Figur 1 verdeutlicht skizzenhaft die Annahme, deren Einführung die Bestimmung der Panning-Koeffizienten 310 ($a_L(b,k)$) und 320 ($a_R(b,k)$) deutlich vereinfacht. In Zeit-Frequenz-Darstellung wird die Zeit im Folgenden grundsätzlich als Blocknummer b des bei der Kurzzeit-Fourier-Transformation (STFT) erhaltenen Blocks angegeben. Das Frequenzband bzw. der Frequenzindex wird mit k indiziert.

[0057] Das Stereoaudiosignal umfasst ein erstes Audiosignal 110 für eine linke Wiedergabeeinrichtung 810 und ein zweites Audiosignal 120 für eine rechte Wiedergabeeinrichtung 820. Durch Kurzzeit-Fourier-Transformation (STFT) wird das erste Audiosignal 110 in seine Zeit-Frequenz-Darstellung 115 ($X_L(b,k)$) überführt. Ebenso wird das zweite Audiosignal 120 in seine Zeit-Frequenz-Darstellung 125 ($X_R(b,k)$) überführt.

[0058] Der Hörer ist an der Position 1 am Rand des Hörbereichs 890 angeordnet. Das durch den Hörer 1, die linke Wiedergabeeinrichtung 810 und die rechte Wiedergabeeinrichtung 820 definierte gleichseitige Dreieck trägt das Bezugszeichen 891 und ist in den kreisförmigen Hörbereich 890 einbeschrieben. Für die Ermittlung der Panning-Koeffizienten 310 und 320 wird nun gemäß der Erfindung angenommen, dass sich eine einzige Direktschallquelle 813, deren Lautstärke 330 in Abhängigkeit der Zeit b und der Frequenz k variiert, entlang des durchgezogenen Kreisbogens 892 am Rand des Hörbereichs 890 im Bereich zwischen der linken Wiedergabeeinrichtung 810 und der rechten Wiedergabeeinrichtung 820 bewegt. Diese Bewegung ist ebenfalls von der Zeit b und von der Frequenz k abhängig. Die aktuelle azimutale Position $\varphi(b,k)$ der Direktschallquelle 813 auf dem Kreisbogen bestimmt die Panning-Koeffizienten 310 und 320. Die komplexe Amplitude 330 der Direktschallquelle 813 ergibt, wenn man sie multiplikativ mit den ersten Panning-Koeffizienten 310 gewichtet, die Zeit-Frequenz-Darstellung 115 des ersten Audiosignals 110. Wird die Signalstärke 330 dagegen mit dem zweiten Panning-Koeffizienten 320 multiplikativ gewichtet, erhält man die Zeit-Frequenz-Darstellung 125 des zweiten Audiosignals 120.

[0059] Figur 2 verdeutlicht den Zusammenhang zwischen dem ersten und zweiten Panning-Koeffizienten 310 und 320 einerseits und dem Positionskoeffizienten 390 (Ψ) andererseits. Aufgetragen ist jeweils der Wert dieser Koeffizienten über der Azimutposition φ von links L über Mitte M nach rechts R. Die Panning-Koeffizienten 310 und 320 verlaufen in Abhängigkeit der Azimut-Position φ nicht linear. Der Positionskoeffizient 390 hat demgegenüber den Vorteil, dass er von links L über Mitte M nach rechts R durchgehend linear verläuft.

[0060] Figur 3 verdeutlicht das Repanning zwecks Wiedergabe des Stereoaudiosignals als Mehrkanalaudiosignal. Das Signal 330 der Direktschallquelle wird mit Repanning-Koeffizienten 410 (g_1), 420 (g_2) und 430 (g_3) gewichtet auf Tonkanäle 580, 585 und 590 übertragen, die auf den drei Lautsprechern L, C und R wiedergegeben werden. In die Ermittlung der Repanning-Koeffizienten 410, 420 und 430 gehen die bei der Analyse des Stereosignals ermittelten Panning-Koeffizienten 310 und 320 ein. Die bei der Analyse weiterhin ermittelten Umgebungssignale 510 und 520 werden zum Einen den Tonkanälen 580 und 590 additiv überlagert. Zum Anderen werden sie auf zusätzlichen Lautsprechern RL und RR wiedergegeben. Alle Lautsprecher L, C, R, RL und RR sind auf einem Kreis K angeordnet, der gleichzeitig den Hörbereich 890 um den Hörer 1 definiert. Die Winkelpositionen der Lautsprecher L, C und R liegen jeweils um 30 Grad auseinander. Die Winkelpositionen der Lautsprecher RL und C bzw. RR und C liegen jeweils um 115 Grad auseinander.

[0061] Figur 4 verdeutlicht skizzenhaft den alternativen Zugang zu den Panning-Koeffizienten 310 und 320 über Gleichungen in Leistungen. Die beiden Audiosignale 110 und 120 werden in diesem Beispiel jeweils mit einer Filterbank 150 im Zeitbereich zerlegt. Die in Figur 4 beispielhaft gezeichnete Filterbank 150 enthält vier Bandpassfilter, die mit Bandindizes $k=1$, $k=2$, $k=3$ und $k=4$ gekennzeichnet sind. Der Filter mit Bandindex $k=1$ lässt nur Frequenzen ω mit $0 < \omega \leq \omega_1$ passieren. Der Filter mit Bandindex $k=2$ lässt nur Frequenzen ω mit $\omega_1 < \omega \leq \omega_2$ passieren. Der Filter mit Bandindex $k=3$ lässt nur Frequenzen ω mit $\omega_2 < \omega \leq \omega_3$ passieren. Schließlich lässt der Filter mit Bandindex $k=4$ nur Frequenzen ω mit $\omega_3 < \omega \leq \omega_4$ passieren.

[0062] Das Ausgangssignal eines jeden Filters ist nach wie vor ein zeitabhängiges Signal. In der Information, aus welchem Filter das Signal kommt, also zu welchem Bandindex k es gehört, steckt die Frequenzinformation. Alle Ausgangssignale $x_{L,R}(b,k=1-4)$ bilden somit zusammen eine Zeit-Frequenz-Darstellung 115 bzw. 125 der Audiosignale 110 bzw. 120. Aus den Ausgangssignalen $x_{L,R}(b,k=1-4)$ wird im Schritt 145 jeweils über eine rekursive Mittelung die zugehörige momentane Leistung $P_{L,R}(b,k=1-4)$ ermittelt. Diese Funktionen bilden gemeinsam die mit dem Bezugszeichen 115a bzw. 125a bezeichnete zeit- und frequenzabhängige Leistung $P_{L,R}(b,k)$ des linken Audiosignals 110 bzw. des rechten Audiosignals 120. Diese Leistung steht auf der linken Seite der Gleichung.

[0063] Auf der rechten Seite der Gleichung steht das Produkt aus dem Quadrat des mit dem Bezugszeichen 310 bzw. 320 bezeichneten Panning-Koeffizienten $a_{L,R}(b,k)$ mit der Leistung $P_s(b,k)$ (Bezugszeichen 330a) des gesuchten Direktsignals $s(b,k)$ (Bezugszeichen 330).

[0064] Figur 5 ist an Figur 1 angelehnt und verdeutlicht skizzenhaft, wie aus den Panning-Koeffizienten 310 ($a_L(b,k)$) und 320 ($a_R(b,k)$) im nächsten Schritt das Direktsignal 330 ($S(b,k)$) sowie die beiden Umgebungssignale 510 ($N_L(b,k)$) und 520 ($N_R(b,k)$) ermittelt werden können. Die Zeit-Frequenz-Darstellung 115 ($X_L(b,k)$) des ersten Audiosignals 110

geht über eine erste Gleichung eindeutig aus dem gesuchten ersten Umgebungssignal 510, dem ebenfalls gesuchten Direktsignal 330 und dem bekannten ersten Panning-Koeffizienten 310 hervor. Ebenso geht die Zeit-Frequenz-Darstellung 125 ($X_R(b,k)$) des zweiten Audiosignals 120 über eine zweite Gleichung eindeutig aus dem gesuchten zweiten Umgebungssignal 520, dem gesuchten Direktsignal 330 und dem bekannten zweiten Panning-Koeffizienten 320 hervor. Diese beiden Gleichungen enthalten drei unbekannte Größen. Um eine eindeutige Lösung zu erhalten, wird eine der Unbekannten eliminiert.

[0065] Zu diesem Zweck wird ausgenutzt, dass beide Umgebungssignale 510 und 520 ähnlich klingen. Es wird daher angenommen, dass sie auf das gleiche gemeinsame Umgebungssignal 530 ($N(b,k)$) zurückgehen, welches lediglich mit zwei unterschiedlichen Dekorrelationsfunktionen 540 ($H_L(k)$) und 550 ($H_R(k)$) gefiltert wurde. Die Dekorrelationsfunktionen 540 und 550 sind zwar nicht bekannt, lassen sich aber nach Stand der Technik beispielsweise als Filterfunktionen mit zufälligem Frequenzgang darstellen. Diese Näherung ist ausreichend, um die beiden Gleichungen nach dem Direktsignal 330 und dem gemeinsamen Umgebungssignal 530 auflösen zu können.

[0066] Im Folgenden wird ein Ausführungsbeispiel des erfindungsgemäßen Verfahrens mathematisch erläutert: Der Verarbeitung liegt ein Signalmodell zu Grunde, welches das in einem Stereoaudiosignal enthaltene, an diskreten Zeitpunkten n aufgenommene erste Audiosignal 110 ($x_L(n)$) für die linke Wiedergabeeinrichtung 810 bzw. das zweite Audiosignal 120 ($x_R(n)$) für die rechte Wiedergabeeinrichtung 820

$$x_L(n) = \left[\sum_{j=1}^J a_{L,j} \cdot s_j(n) \right] + n_L(n) = a_{L,1} \cdot s_1(n) + a_{L,2} \cdot s_2(n) + \dots + n_L(n) \quad (1)$$

$$x_R(n) = \left[\sum_{j=1}^J a_{R,j} \cdot s_j(n) \right] + n_R(n) = a_{R,1} \cdot s_1(n) + a_{R,2} \cdot s_2(n) + \dots + n_R(n) \quad (2)$$

als die gewichtete Summe einzelner Quellsignale $s_j(n)$ beschreibt, wobei $j=1, \dots, J$ die einzelnen Schallquellen indiziert. Der linke Kanal x_L und der rechte Kanal x_R enthalten außerdem die jeweils nicht richtungsabhängigen, diffusen Umgebungssignale $n_L(n)$ bzw. $n_R(n)$. Die Panning-Koeffizienten $a_{L,j}$ bzw. $a_{R,j}$ geben jeweils eine richtungsabhängige Gewichtung an, mit der die nur von der Zeit abhängigen Quellsignale $s_j(n)$ in das erste Audiosignal x_L bzw. in das zweite Audiosignal x_R eingehen.

[0067] Die Panning-Koeffizienten $a_{L,j}$ und $a_{R,j}$ können über die Beziehung $a_{L,j}^2 + a_{R,j}^2 = 1$ miteinander verknüpft werden, was dazu führt, dass unabhängig von der Position der einzelnen Quellen eine konstante Lautheit erzielt wird. Dies entspricht dem üblicherweise in der Musikproduktion verwendeten Constant-Power-Panning.

[0068] Die Signale können nun auf verschiedene Weise in eine Zeit-Frequenz-Darstellung überführt werden. Beispielsweise kann eine Kurzzeit-Fouriertransformation (STFT) durchgeführt werden. Eine Zeit-Frequenz-Darstellung kann aber auch unmittelbar aus den zeitabhängigen Signalen erhalten werden. Beispielsweise können die Signale mit einer Filterbank, die aus mehreren parallel geschalteten Bandpassfiltern besteht, in Anteile zerlegt werden, die ein jeder dieser Bandpassfilter durchlässt. Jeder dieser Anteile ist dann nach wie vor ein zeitabhängiges Signal. Unabhängig davon, auf welchem Wege die Zeit-Frequenz-Darstellung erhalten wurde, lässt sie sich als

$$X_L(b,k) = \sum_{j=1}^J a_{L,j} \cdot S_j(b,k) + N_L(b,k) \quad (3)$$

$$X_R(b,k) = \sum_{j=1}^J a_{R,j} \cdot S_j(b,k) + N_R(b,k) \quad (4)$$

schreiben. Wurde die Zeit-Frequenz-Darstellung durch Kurzzeit-Fouriertransformation (STFT) erhalten, bezeichnet man b üblicherweise als Blockindex und k als Frequenzindex. Wurde die Zeit-Frequenz-Darstellung hingegen unmittelbar aus den zeitabhängigen Signalen erhalten, beispielsweise mit einer Filterbank, bezeichnet man b üblicherweise als Zeitindex und k als Bandindex, da die Diskretisierung der Frequenzen durch die jeweils von den Bandpassfiltern durchgelassenen Frequenzbänder bestimmt wird.

[0069] Die Koeffizienten $a_{R,j}$ und $a_{L,j}$ können weiterhin zu einem Positions-Koeffizienten

$$\Psi_j = a_{R,j}^2 - a_{L,j}^2 \quad (5)$$

zusammengefasst werden. Dieser steht in einem linearen Zusammenhang zur Azimut-Position, wobei der Wertebereich von [-1,...,1] sich auf maximal links bzw. maximal rechts gepannte Signale abbildet (Figur 2). Dies erlaubt eine intuitive Zuordnung zwischen dem Wert des Koeffizienten und der tatsächlichen Position im Stereopanorama.

[0070] Wenn statt der Amplituden $X_L(b,k)$ und $X_R(b,k)$ die Leistungen $P_L(b,k)$ und $P_R(b,k)$ miteinander verglichen werden, ist es zweckmäßiger, den Positions-Koeffizienten als

$$\Psi_j = \frac{a_{R,j} - a_{L,j}}{a_{R,j} + a_{L,j}} \quad (5a)$$

zu schreiben. Er steht dann nach wie vor in dem in Figur 2 gezeigten linearen Zusammenhang zur Azimut-Position.

[0071] Unter der Annahme, dass in den Gleichungen (3) und (4) in einem Frequenzband k jeweils nur eine dominante Quelle auftritt, können die einzelnen Quellen $S_j(b,k)$ zu einer einzigen ungepannten Mischquelle (Direktschallquelle) mit einer zeit- und frequenzabhängigen komplexen Amplitude $S(b,k) = \sum S_j(b,k)$ zusammengeführt werden. Die Auswirkung dieser Mischquelle auf die Signale $X_L(b,k)$ bzw. $X_R(b,k)$ ist dann ebenfalls zeit- und frequenzabhängig und wird durch die Panning-Koeffizienten $a_L(b,k)$ bzw. $a_R(b,k)$ beschrieben:

$$X_L(b,k) = a_L(b,k) \cdot S(b,k) + N_L(b,k) \quad (3a)$$

$$X_R(b,k) = a_R(b,k) \cdot S(b,k) + N_R(b,k) \quad (4a)$$

[0072] Unter Vernachlässigung der, im Vergleich zu S üblicherweise relativ kleinen, diffusen Umgebungssignale N_L bzw. N_R ergibt sich insgesamt folgendes Gleichungssystem für die Panning-Koeffizienten $a_L(b,k)$ und $a_R(b,k)$:

$$a_L^2(b,k) + a_R^2(b,k) = 1 \quad (6)$$

$$X_L(b,k) = a_L(b,k) \cdot S(b,k) \quad (7)$$

$$X_R(b,k) = a_R(b,k) \cdot S(b,k) \quad (8)$$

[0073] Durch Auflösen erhält man die Panning-Koeffizienten

$$a_L(b,k) = \sqrt{\frac{X_L(b,k)^2}{X_L(b,k)^2 + X_R(b,k)^2}} \quad (9)$$

$$a_R(b,k) = \sqrt{\frac{X_R(b,k)^2}{X_L(b,k)^2 + X_R(b,k)^2}} \quad (10)$$

[0074] Die Signale x_L , x_R und S sind im Allgemeinen komplexwertig, während die Panning-Koeffizienten a_L und a_R reellwertig sind, da in dem Signalmodell gemäß den Gleichungen (7) und (8) ein reines Amplitudenpanning durchgeführt wird, d.h. nur die Amplitude richtungsabhängig ist. Daraus folgt, dass sowohl $X_L(b,k)$ als auch $X_R(b,k)$ in Phase mit $S(b,k)$ sind. In den Polardarstellungen

$$a_L(b, k) = \sqrt{\frac{|X_L(b, k)|^2 \cdot \exp(-2i\varphi_L)}{|X_R(b, k)|^2 \cdot \exp(-2i\varphi_R) + |X_L(b, k)|^2 \cdot \exp(-2i\varphi_L)}} \quad (11)$$

$$a_R(b, k) = \sqrt{\frac{|X_R(b, k)|^2 \cdot \exp(-2i\varphi_R)}{|X_R(b, k)|^2 \cdot \exp(-2i\varphi_R) + |X_L(b, k)|^2 \cdot \exp(-2i\varphi_L)}} \quad (12)$$

sind also die Phasen ϕ_L von X_L , ϕ_R von X_R und ϕ_S von S identisch, so dass sich die Phasenterme kürzen lassen:

$$a_L(b, k) = \sqrt{\frac{|X_L(b, k)|^2}{|X_R(b, k)|^2 + |X_L(b, k)|^2}} \quad (13)$$

$$a_R(b, k) = \sqrt{\frac{|X_R(b, k)|^2}{|X_R(b, k)|^2 + |X_L(b, k)|^2}} \quad (14)$$

[0075] Die Panning-Koeffizienten a_L und a_R sind in dieser Näherung also unmittelbar mit den Leistungsdichtespektren (Zeit-Frequenz-Darstellungen) X_L und X_R des ersten und zweiten Audiosignals, die zusammen das Stereoaudiosignal ergeben, verknüpft.

[0076] Alternativ kann je nach Bedarf und Anwendung auch der Positionskoeffizient

$$\Psi(b, k) = \frac{|X_R(b, k)|^2 - |X_L(b, k)|^2}{|X_R(b, k)|^2 + |X_L(b, k)|^2} \quad (15)$$

berechnet werden. Dieser Positionskoeffizient $\Psi(b, k)$ erlaubt eine sehr effektive Berechnung der Position durch einfache Betrachtung des Differenzleistungsspektrums und der Gesamtleistung des Signals.

[0077] Da mit dem Kanalmodell (7-8) ein reines Amplitudenpanning durchgeführt wird, folgt, dass der linke und rechte Kanal (X_L und X_R) in Phase mit dem Direktsignal S sind. Somit lässt sich das Kanalmodell auch über die Leistungen ausdrücken:

$$P_L(b, k) = a_L^2(b, k) \cdot P_S(b, k) \quad (7a)$$

$$P_R(b, k) = a_R^2(b, k) \cdot P_S(b, k). \quad (8a)$$

[0078] Hierin ist $P_L(b, k)$ die Leistung des linken Kanals X_L , $P_R(b, k)$ ist die Leistung des rechten Kanals X_R , und P_S ist die Leistung des Direktsignals S .

[0079] Wurde die Zeit-Frequenz-Darstellung durch Kurzzeit-Fourier-Transformation (STFT) erhalten, entspricht eine Leistung $P_X(b, k)$ dem Leistungsdichtespektrum $|X(b, k)|^2$.

[0080] Wurde die Zeit-Frequenz-Darstellung hingegen beispielsweise durch Filterbankzerlegung im Zeitbereich erhalten, existiert nicht notwendigerweise eine geschlossene Formel für die momentane Leistung $P_x(b, k)$ pro Band k . Diese momentane Leistung lässt sich jedoch beispielsweise über rekursive Mittelung

$$P_x(b, k) = \alpha \cdot P_x(b-1, k) + (1-\alpha) \cdot [x(b, k)]^2, \quad 0 < \alpha < 1 \quad (8b)$$

gewinnen. Der Kleinbuchstabe x symbolisiert, dass die Zeit-Frequenz-Darstellung $x(b, k)$ durch Zerlegung im Zeitbereich erhalten wurde.

[0081] Das Quadrat des momentanen Signals wird also als Maß dafür gewertet, wie stark sich die momentane Leistung

$P_x(b,k)$ zum Zeitpunkt b gegenüber dem vorherigen Zeitpunkt $b-1$ ändert, α ist ein Gewichtungsfaktor, mit dem das Festhalten am bisherigen Trend für die momentane Leistung $P_x(b,k)$ gegen die Berücksichtigung neuer Informationen abgewogen wird. Dieser ist vorzugsweise so klein zu wählen, dass eine stabile Schätzung der mittleren Leistung erfolgt, ohne dass es durch Transienten oder kurzzeitige Signaländerungen zu starken Schwankungen kommt.

[0082] Durch Auflösen von (7a) und (8a) erhält man die Panning-Koeffizienten

$$a_L(b,k) = \sqrt{\frac{P_L(b,k)}{P_L(b,k) + P_R(b,k)}} \quad (9a)$$

$$a_R(b,k) = \sqrt{\frac{P_R(b,k)}{P_L(b,k) + P_R(b,k)}} \quad (10a)$$

[0083] Alternativ kann je nach Bedarf und Anwendung auch der Positions-Koeffizient

$$\Psi(b,k) = \frac{P_R(b,k) - P_L(b,k)}{P_R(b,k) + P_L(b,k)} \quad (15a)$$

berechnet werden. Optional kann hierbei eine weitere Anpassung an das Gehör erfolgen, indem in Gleichung (15a) die Leistungen $P_L(b,k)$ und $P_R(b,k)$ jeweils durch ihre Wurzeln ersetzt werden. Der Positions-Koeffizient $\Psi(b,k)$ vermittelt dann einen noch realistischeren Eindruck von der Position der Direktschallquelle.

[0084] Auf Grund der vereinfachenden Annahmen, unter denen die Panning-Koeffizienten a_L und a_R sowie die Position Ψ gewonnen wurden, sind diese Größen Näherungswerte. Sie werden im Folgenden zur Unterscheidung von den exakten Werten gemäß dem Signalmodell mit $\hat{a}_L$, $\hat{a}_R$ bzw. $\hat{\Psi}$ bezeichnet.

[0085] Zur Extraktion des Direktsignals S und der Umgebungssignale N_L und N_R aus den Summensignalen X_L und X_R (Gleichungen (3) und (4)) kommt ein iteratives Verfahren zum Einsatz. Aus dem linken Eingangskanal X_L und dem rechten Eingangskanal X_R werden schrittweise Direktsignalbeiträge $\hat{S}_i$ extrahiert, die am Ende zum Direktsignal $\hat{S}$ der Direktschallquelle zusammengefasst werden. Die Differenz zwischen dem mit den Panning-Koeffizienten a_L und a_R gewichteten Direktsignal $\hat{S}$ und den Eingangssignalen X_L bzw. X_R ist eine Näherung für die Umgebungssignale N_L bzw. N_R . Die Indices (b,k) werden im Folgenden auf Grund der besseren Übersichtlichkeit nicht mehr explizit angegeben.

[0086] Zum Start der Iteration werden die geschätzten Umgebungssignale $\hat{N}_L$ und $\hat{N}_R$ zunächst mit den Eingangssignalen X_L und X_R initialisiert:

$$\hat{N}_{L,0} = X_L, \quad \hat{N}_{R,0} = X_R \quad (16)$$

[0087] Ausgehend hiervon werden gemäß den Iterationsvorschriften

$$\hat{a}_{L,i} = \sqrt{\frac{|\hat{N}_{L,i-1}|^2}{|\hat{N}_{R,i-1}|^2 + |\hat{N}_{L,i-1}|^2}} \quad (17)$$

$$\hat{a}_{R,i} = \sqrt{\frac{|\hat{N}_{R,i-1}|^2}{|\hat{N}_{R,i-1}|^2 + |\hat{N}_{L,i-1}|^2}} \quad (18)$$

$$\hat{S}_i = \frac{(\hat{N}_{L,i-1} + \hat{N}_{R,i-1})}{2} \quad (19)$$

die Panning-Koeffizienten verfeinert und ein Direktsignalbeitrag berechnet. Bei der ersten Iteration haben die Panning-Koeffizienten genau die Werte gemäß den Gleichungen (13) und (14) als Startwerte. Die Berechnung des Direktsignalbeitrags $\hat{S}_i$ gemäß Gleichung (19) geht davon aus, dass das Direktsignal im ersten und zweiten Audiosignal mit gleicher Phase vorliegt und die Umgebungssignale dazu phasenverschoben sind.

[0088] Vor der nächsten Iteration werden die Umgebungssignale über

$$\hat{N}_{L,i} = \hat{N}_{L,i-1} - \hat{a}_{L,i} \cdot \hat{S}_i \quad (20)$$

$$\hat{N}_{R,i} = \hat{N}_{R,i-1} - \hat{a}_{R,i} \cdot \hat{S}_i \quad (21)$$

in dem Sinne selbstkonsistent nachgeführt, dass ein Signalanteil, der sich als ein mit der Direktschallquelle 813 korrelierter Direktsignalanteil erwiesen hat, nicht gleichzeitig zum diffusen Umgebungssignal gehören kann. Diese selbstkonsistente Lösung zeichnet sich insbesondere dadurch aus, dass sie eine gute Extraktion stark gepannter, d.h. stark richtungsabhängiger, Direktsignale ermöglicht.

[0089] Nach Durchlauf aller I Iterationen ergibt sich das gesamte, mit der Direktschallquelle 813 korrelierte Direktsignal als Summe der einzelnen Signalanteile $\hat{S}_i$:

$$\hat{S} = \sum_{i=1}^I \hat{S}_i \quad (22)$$

[0090] Bei der Ermittlung der Panning-Koeffizienten $\hat{a}_{L,i}$ und $\hat{a}_{R,i}$ sowie der Signalanteile $\hat{S}_i$ wurde ausschließlich Selbstkonsistenz mit den Umgebungssignalen $\hat{N}_{L,i}$ und $\hat{N}_{R,i}$ gefordert, ohne dass das Signalmodell gemäß den Gleichungen (3) und (4) herangezogen wurde. Daher ist nicht sichergestellt, dass die letztendlich erhaltenen Werte für $\hat{N}_L$, $\hat{N}_R$ und $\hat{S}$ diesem Signalmodell gehorchen. Da sich eine Verletzung des Signalmodells stärker auf den Höreindruck auswirkt als eine Abweichung im diffusen Umgebungssignal, wird der Erfüllung des Signalmodells Priorität gegenüber einer möglichst exakten Näherung für $\hat{N}_L$ und $\hat{N}_R$ eingeräumt. Daher werden nicht die bei der letzten Iteration I erhaltenen Werte $\hat{N}_{L,I}$ und $\hat{N}_{R,I}$ als Umgebungssignale $\hat{N}_L$ und $\hat{N}_R$ verwendet, sondern diese werden am Ende aus dem Gesamtergebnis $\hat{S}$ für das Direktsignal und den ersten Näherungswerten $\hat{a}_{L,1}$ und $\hat{a}_{R,1}$ für die Panning-Koeffizienten berechnet:

$$\hat{N}_L = X_L - \hat{a}_{L,1} \cdot \hat{S} \quad (23)$$

$$\hat{N}_R = X_R - \hat{a}_{R,1} \cdot \hat{S} \quad (24)$$

[0091] Die während des iterativen Verfahrens gemäß den Gleichungen (17) und (18) verfeinerten Panning-Koeffizienten werden ausschließlich für die Aufteilung der Signale X_L und X_R in Direktsignal $\hat{S}$ und Umgebungssignale $\hat{N}_L$ und $\hat{N}_R$ verwendet. Für das Repanning auf eine Konfiguration von mehr als zwei Lautsprechern werden weiterhin die aus der Lösung des Gleichungssystems (13-14) erhaltenen Panning-Koeffizienten verwendet.

[0092] Für $i \rightarrow \infty$ gilt für die Umgebungssignale $\hat{N}_{L,i}$ und $\hat{N}_{R,i}$ gemäß den Gleichungen (20) und (21)

$$\hat{N}_{L,i} = -\hat{N}_{R,i} \quad (25)$$

[0093] Es sind also beide Umgebungssignale bis auf eine Phasendrehung identisch. Das ursprüngliche Signalmodell gemäß den Gleichungen (3a) und (4a) vereinfacht sich damit zu

$$X_L = a_L \cdot S + N \quad (26)$$

$$X_R = a_R \cdot S - N \quad (27)$$

[0094] Einsetzen der Panning-Koeffizienten gemäß den Gleichungen (13) und (14) sowie Auflösen ergibt

$$\hat{S} = \frac{X_L + X_R}{\hat{a}_L + \hat{a}_R} \quad (28)$$

$$\hat{N} = \frac{\hat{a}_R \cdot X_L - \hat{a}_L \cdot X_R}{\hat{a}_L + \hat{a}_R}$$

als Näherungswerte für das Direktsignal $\hat{S}$ und das Umgebungssignal

$$\hat{N}_L \equiv -\hat{N}_R \equiv \hat{N}.$$

[0095] Im Folgenden wird ein allgemeinerer Ansatz zur Bestimmung des Direktsignals und der Umgebungssignale aus den Panning-Koeffizienten gegeben. Dieser Ansatz geht von der Annahme aus, dass beide Umgebungssignale ähnlich klingen, jedoch durch unterschiedliche Ausbreitungswege und Reflexionen dekorreliert sind.

[0096] Somit lassen sich beide Umgebungssignale $\hat{N}_L$ und $\hat{N}_R$ als Filterungen eines gemeinsamen Umgebungssignals N mit unterschiedlichen Dekorrelationsfunktionen H_L und H_R darstellen:

$$\hat{N}_L(b, k) = H_L \{N(b, k)\}, \quad (29)$$

$$\hat{N}_R(b, k) = H_R \{N(b, k)\}. \quad (30)$$

[0097] Eine Filterung lässt sich in einer Zeit-Frequenz-Darstellung als bandweise Multiplikation mit einem Verstärkungsfaktor und einer Phasendrehung ausdrücken. $X_L(b, k)$ und $X_R(b, k)$ sind dann über die beiden Gleichungen

$$X_L(b, k) = a_L(b, k) \cdot S(b, k) + H_L(b, k) \cdot N(b, k), \quad (31)$$

$$X_R(b, k) = a_R(b, k) \cdot S(b, k) + H_R(b, k) \cdot N(b, k) \quad (32)$$

mit dem Direktsignal S und dem Umgebungssignal N verknüpft.

[0098] Die allgemeine Form der Dekorrelationsfunktionen $H_{L,R}(b, k)$ kann, sofern die Zeit-Frequenz-Darstellungen $X_L(b, k)$ und $X_R(b, k)$ aus einer vollständigen Transformation in den Frequenzbereich gewonnen wurden, etwa mittels Kurzzeit-Fourier-Transformation (STFT), als komplexes Spektrum

$$H_{L,R}(k) = \gamma(k) \cdot \exp(i\phi(k)), \quad 0 < \gamma(k) < 1, \quad 0 < \phi(k) < \pi \quad (33)$$

mit einer frequenzabhängigen Amplitude $\gamma(k)$ und Phase $\phi(k)$, beschrieben werden.

[0099] Einsetzen der Panning-Koeffizienten aus den Gleichungen (9) und (10) in die Gleichungen (31) und (32) und Auflösen ergibt

$$\hat{S}(b, k) = \frac{X_L(b, k) \cdot H_R(k) - X_R(b, k) \cdot H_L(k)}{\hat{a}_L(b, k) \cdot H_R(k) - \hat{a}_R(b, k) \cdot H_L(k)}, \quad (34)$$

$$\hat{N}(b, k) = \frac{\hat{a}_L(b, k) \cdot X_R(b, k) - \hat{a}_R(b, k) \cdot X_L(b, k)}{\hat{a}_L(b, k) \cdot H_R(k) - \hat{a}_R(b, k) \cdot H_L(k)} \quad (35)$$

für das geschätzte Direktsignal S und für das gemeinsame Umgebungssignal $\hat{N}$.

[0100] Werden Zeit-Frequenz-Darstellungen $x_L(b, k)$ und $x_R(b, k)$ mit einer Filterbank gewonnen, werden die Gleichungen (31) und (32) zu

$$x_L(b, k) = a_L(b, k) \cdot s(b, k) + h_L \{n(b, k)\}, \quad (36)$$

$$x_R(b, k) = a_R(b, k) \cdot s(b, k) + h_R \{n(b, k)\}, \quad (37)$$

worin die Benennung von h , x , a , s und n in Kleinbuchstaben wiederum verdeutlicht, dass es sich um Größen im Zeitbereich handelt. Die Dekorrelationsfunktionen H_L und H_R lassen sich jetzt nicht mehr so einfach anwenden wie im Frequenzbereich. Mit der Einschränkung

$$h_{L,R}(k) = \gamma(k) \cdot (\pm 1), \quad (38)$$

wonach die Dekorrelationsfunktion pro Band nur Phasenverschiebungen von 0 (+1) und π (-1) erzeugen kann, vereinfachen sich die Gleichungen (36) und (37) zu

$$x_L(b, k) = a_L(b, k) \cdot s(b, k) + h_L(k) \cdot n(b, k), \quad (39)$$

$$x_R(b, k) = a_R(b, k) \cdot s(b, k) + h_R(k) \cdot n(b, k). \quad (40)$$

[0101] Nach mathematischem Umformen ergeben sich

$$\hat{s}(b, k) = \frac{h_R(k) \cdot x_L(b, k) - h_L(k) \cdot x_R(b, k)}{h_R(k) \cdot \hat{a}_L(b, k) - h_L(k) \cdot \hat{a}_R(b, k)}, \quad (41)$$

$$\hat{n}(b, k) = \frac{\hat{a}_L(b, k) \cdot x_R(b, k) - \hat{a}_R(b, k) \cdot x_L(b, k)}{\hat{a}_L(b, k) \cdot h_R(k) - \hat{a}_R(b, k) \cdot h_L(k)} \quad (42)$$

als die Lösungen für die Direkt- und Umgebungssignale.

Bezugszeichenliste

[0102]

1	Position des Hörers
110	erstes (linkes) Audiosignal x_L des Stereoaudiosignals
115	Zeit-Frequenz-Darstellung X_L des ersten Audiosignals 110
115a	zeit- und frequenzabhängige Leistung P_L des Signals 110
120	zweites (rechtes) Audiosignal x_R des Stereoaudiosignals
125	Zeit-Frequenz-Darstellung X_R des zweiten Audiosignals 120
125a	zeit- und frequenzabhängige Leistung P_R des Signals 120
145	Ermittlung der zeit- und frequenzabhängigen Leistung 115a, 125a
150	Filterbank
310	Panning-Koeffizienten $a_L(b, k)$ des ersten Audiosignals 110
320	Panning-Koeffizienten $a_R(b, k)$ des zweiten Audiosignals 120
330	komplexe Amplitude $S(b, k)$ der Direktschallquelle 813
330a	zeit- und frequenzabhängige Leistung P_S des Signals 330
φ	azimutale Position der Direktschallquelle 813
390	Positionskoeffizient Ψ
410	erster Repanning-Koeffizient g_1 für ersten Tonkanal 580
420	zweiter Repanning-Koeffizient g_2 für zweiten Tonkanal 585
430	dritter Repanning-Koeffizient g_3 für dritten Tonkanal 590
510	erstes (linkes) Umgebungssignal N_L
520	zweites (rechtes) Umgebungssignal N_R

530	gemeinsames Umgebungssignal $N(b,k)$
540	erste Dekorrelationsfunktion $H_L(k)$
550	zweite Dekorrelationsfunktion $H_R(k)$
580	erster Tonkanal für Lautsprecher auf Position L (links)
585	zweiter Tonkanal für Lautsprecher auf Position C (Mitte)
590	dritter Tonkanal für Lautsprecher auf Position R (rechts)
810	linke Wiedergabeeinrichtung für das erste Audiosignal 110
813	Direktschallquelle
820	rechte Wiedergabeeinrichtung für das zweite Audiosignal 120
890	Hörbereich vor dem Hörer 1 bzw. um den Hörer 1
891	gleichseitiges Dreieck im Hörbereich 890
892	Kreisbogen am Rand des Hörbereichs 890
L, C, R	Lautsprecherpositionen Links, Mitte, Rechts für das Repanning
RL, RR	zusätzliche Lautsprecherpositionen für Umgebungssignale 510, 520

Patentansprüche

- Verfahren zur Analyse eines Stereoaudiosignals, wobei das Stereoaudiosignal ein erstes Audiosignal (110) für eine linke Wiedergabeeinrichtung (810) und ein zweites Audiosignal (120) für eine rechte Wiedergabeeinrichtung (820) aufweist, mit folgende Schritten:

das erste Audiosignal (110) wird in eine erste Zeit-Frequenz-Darstellung (115) überführt, und das zweite Audiosignal (120) wird in eine zweite Zeit-Frequenz-Darstellung (125) überführt;

aus der ersten Zeit-Frequenz-Darstellung (115) wird die zeit- und frequenzabhängige Leistung (115a) des ersten Audiosignals (110) ermittelt (145), und aus der zweiten Zeit-Frequenz-Darstellung (125) wird die zeit- und frequenzabhängige Leistung (125a) des zweiten Audiosignals (120) ermittelt (145);

es wird eine erste Gleichung aufgestellt, die die zeit- und frequenzabhängige Leistung (115a) des ersten Audiosignals (110) in Beziehung setzt zum Produkt des Quadrats eines ersten zeit- und frequenzabhängigen Panning-Koeffizienten (310) mit der zeit- und frequenzabhängigen Leistung (330a) einer in einem Hörbereich (890) zwischen der linken Wiedergabeeinrichtung (810) und der rechten Wiedergabeeinrichtung (820) angeordneten Direktschallquelle (813);

es wird eine zweite Gleichung aufgestellt, die die zeit- und frequenzabhängige Leistung (125a) des zweiten Audiosignals (120) in Beziehung setzt zum Produkt des Quadrats eines zweiten zeit- und frequenzabhängigen Panning-Koeffizienten (320) mit der gleichen zeit- und frequenzabhängigen Leistung (330a) der gleichen Direktschallquelle (813);

wobei die Panning-Koeffizienten (310) und (320) dazu ausgebildet sind, die Direktschallquelle (813) in dem Hörbereich (890) zu positionieren;

ein Positionskoeffizient (390), der dem Verhältnis einer Differenz der Panning-Koeffizienten (310) und (320) zur Summe der Panning-Koeffizienten (310) und (320) entspricht, wird als Lösung des aus beiden Gleichungen gebildeten Gleichungssystems ermittelt,

dadurch gekennzeichnet, dass das Gleichungssystem unter der zusätzlichen Bedingung gelöst wird, dass die Summe der Quadrate der Panning-Koeffizienten (310) und (320) konstant ist, wobei der Positionskoeffizient (390) aus dem Verhältnis der Differenz der Wurzeln der zeit- und frequenzabhängigen Leistungen (125a, 115a) beider Audiosignale (110, 120) zur Summe der Wurzeln der zeit- und frequenzabhängigen Leistungen (125a, 115a) beider Audiosignale (110, 120) ermittelt wird.

- Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** der erste Panning-Koeffizient (310) als Wurzel aus dem Verhältnis der zeit- und frequenzabhängigen Leistung (115a) des ersten Audiosignals (110) zur Summe der zeit- und frequenzabhängigen Leistungen (115a, 125a) beider Audiosignale (110, 120) ermittelt wird und dass der zweite Panning-Koeffizient (320) als Wurzel aus dem Verhältnis der zeit- und frequenzabhängigen Leistung (125a) des zweiten Audiosignals (120) zur Summe der zeit- und frequenzabhängigen Leistungen (115a, 125a) beider Audiosignale (110, 120) ermittelt wird.

- Verfahren nach einem der Ansprüche 1 bis 2, **dadurch gekennzeichnet, dass** die zeit- und frequenzabhängige Leistung (115a, 125a) mindestens eines Audiosignals (110, 120) zu einem interessierenden Zeitpunkt als gewichtete Summe aus der zeit- und frequenzabhängigen Leistung (115a, 125a) des Audiosignals zu einem früheren Zeitpunkt und dem Quadrat der Zeit-Frequenz-Darstellung (115, 125) dieses Audiosignals (110, 120) zu dem interessierenden

Zeitpunkt ermittelt wird (145).

4. Verfahren zur Analyse eines Stereoaudiosignals, wobei das Stereoaudiosignal ein erstes Audiosignal (110) für eine linke Wiedergabeeinrichtung (810) und ein zweites Audiosignal (120) für eine rechte Wiedergabeeinrichtung (820) aufweist, mit folgenden Schritten:

das erste Audiosignal (110) wird in eine erste Zeit-Frequenz-Darstellung (115) überführt, und das zweite Audiosignal (120) wird in eine zweite Zeit-Frequenz-Darstellung (125) überführt;

es wird eine erste Gleichung aufgestellt, die die erste Zeit-Frequenz-Darstellung (115) in Beziehung setzt zum Produkt eines ersten zeit- und frequenzabhängigen Panning-Koeffizienten (310) mit dem zeit- und frequenzabhängigen Signal (330) einer in einem Hörbereich (890) zwischen der linken Wiedergabeeinrichtung (810) und der rechten Wiedergabeeinrichtung (820) angeordneten Direktschallquelle (813);

es wird eine zweite Gleichung aufgestellt, die die zweite Zeit-Frequenz-Darstellung (125) in Beziehung setzt zum Produkt eines zweiten zeit- und frequenzabhängigen Panning-Koeffizienten (320) mit dem gleichen Signal (330) der gleichen Direktschallquelle (813);

wobei die Panning-Koeffizienten (310) und (320) dazu ausgebildet sind, die Direktschallquelle (813) in dem Hörbereich (890) zu positionieren;

ein Positionskoeffizient (390), der der Differenz der Quadrate der Panning-Koeffizienten (310) und (320) entspricht, wird als Lösung des aus beiden Gleichungen gebildeten Gleichungssystems ermittelt,

wobei das Gleichungssystem unter der zusätzlichen Bedingung gelöst wird, dass die Summe der Quadrate der Panning-Koeffizienten (310) und (320) konstant ist,

dadurch gekennzeichnet, dass der Positionskoeffizient (390) aus dem Verhältnis der Differenz der Betragsquadrate beider Zeit-Frequenz-Darstellungen (115) und (125) zur Summe der Betragsquadrate beider Zeit-Frequenz-Darstellungen (115) und (125) ermittelt wird.

5. Verfahren nach Anspruch 4, **dadurch gekennzeichnet, dass** der erste Panning-Koeffizient (310) als Wurzel aus dem Verhältnis des Quadrats der Zeit-Frequenz-Darstellung (115) des ersten Audiosignals (110) zur Summe der Quadrate der Zeit-Frequenz-Darstellungen (115) und (125) beider Audiosignale (110) und (120) ermittelt wird und dass der zweite Panning-Koeffizient (320) als Wurzel aus dem Verhältnis der Zeit-Frequenz-Darstellung (125) des zweiten Audiosignals (120) zur Summe der Quadrate der Zeit-Frequenz-Darstellungen (115) und (125) beider Audiosignale (110) und (120) ermittelt wird.

6. Verfahren nach einem der Ansprüche 1 bis 5, **dadurch gekennzeichnet, dass** aus den Panning-Koeffizienten (310) und (320) das Signal (330) der Direktschallquelle (813)

und/oder

zwei nicht mit dieser Direktschallquelle (813) korrelierte Umgebungssignale (510) und (520), wobei das erste Umgebungssignal (510) nur in der Zeit-Frequenz-Darstellung (115) des ersten Audiosignals (110) und das zweite Umgebungssignal (520) nur in der Zeit-Frequenz-Darstellung (125) des zweiten Audiosignals (120) enthalten ist, ermittelt werden.

7. Verfahren nach Anspruch 6, **gekennzeichnet durch** folgende Schritte:

es wird eine erste Gleichung aufgestellt, die die erste Zeit-Frequenz-Darstellung (115) in Beziehung setzt zur Summe aus dem Produkt des ersten Panning-Koeffizienten (310) mit dem zeit- und frequenzabhängigen Signal (330) der Direktschallquelle (813) sowie aus der Filterung eines einzigen gemeinsamen Umgebungssignals (530) mit einer ersten Dekorrelationsfunktion (540);

es wird eine zweite Gleichung aufgestellt, die die zweite Zeit-Frequenz-Darstellung (125) in Beziehung setzt zur Summe aus dem Produkt des zweiten Panning-Koeffizienten (320) mit dem zeit- und frequenzabhängigen Signal (330) der Direktschallquelle (813) sowie aus der Filterung des gemeinsamen Umgebungssignals (530) mit einer zweiten Dekorrelationsfunktion (550);

das zeit- und frequenzabhängige Signal (330) der Direktschallquelle (813), und/oder das gemeinsame Umgebungssignal (530), werden als Lösungen des aus den beiden Gleichungen gebildeten Gleichungssystems ermittelt.

8. Verfahren nach Anspruch 7, **dadurch gekennzeichnet, dass** das zeit- und frequenzabhängige Signal (330) der Direktschallquelle (813) als Differenz zwischen dem frequenzbandweisen Produkt der ersten Zeit-Frequenz-Darstellung (115) mit der zweiten Dekorrelationsfunktion (550) und dem frequenzbandweisen Produkt der zweiten Zeit-Frequenz-Darstellung (125) mit der ersten Dekorrelationsfunktion (540), dividiert durch die Differenz zwischen der

Faltung des ersten Panning-Koeffizienten (310) mit der zweiten Dekorrelationsfunktion (550) und dem frequenzbandweisen Produkt des zweiten Panning-Koeffizienten (320) mit der ersten Dekorrelationsfunktion (540), ermittelt wird.

9. Verfahren nach einem der Ansprüche 7 bis 8, **dadurch gekennzeichnet, dass** das gemeinsame Umgebungssignal (530) als Differenz zwischen dem Produkt der zweiten Zeit-Frequenz-Darstellung (125) mit dem ersten Panning-Koeffizienten (310) und dem Produkt der ersten Zeit-Frequenz-Darstellung (115) mit dem zweiten Panning-Koeffizienten (320), dividiert durch die Differenz zwischen dem frequenzbandweisen Produkt des ersten Panning-Koeffizienten (310) mit der zweiten Dekorrelationsfunktion (550) und dem frequenzbandweisen Produkt des zweiten Panning-Koeffizienten (320) mit der ersten Dekorrelationsfunktion (540), ermittelt wird.
10. Verfahren nach Anspruch 6, **dadurch gekennzeichnet, dass** das Signal (330) der Direktschallquelle (813) und die Umgebungssignale (510, 520) mit einem iterativen Verfahren bestimmt werden ausgehend von einer Iterationsvorschrift, die das Signal (330) der Direktschallquelle einer jeden Iteration, und/oder einen Beitrag zu diesem Signal, in Beziehung setzt zu den Umgebungssignalen (510, 520) der vorherigen Iteration.
11. Verfahren nach Anspruch 10, **dadurch gekennzeichnet, dass** bei jeder Iteration die Panning-Koeffizienten (310) und (320) aus den Umgebungssignalen (510, 520) der vorherigen Iteration neu berechnet werden.
12. Verfahren nach Anspruch 11, **dadurch gekennzeichnet, dass** das erste Umgebungssignal (510) bei jeder Iteration um einen Betrag korrigiert wird, der das Produkt aus dem neu berechneten ersten Panning-Koeffizienten (310) mit dem Signal (330) der Direktschallquelle (813) gemäß der aktuellen Iteration ist, und dass das zweite Umgebungssignal (520) bei jeder Iteration um einen Betrag korrigiert wird, der das Produkt aus dem neu berechneten zweiten Panning-Koeffizienten (320) mit dem Signal (330) der Direktschallquelle (813) gemäß der aktuellen Iteration ist.
13. Verfahren nach einem der Ansprüche 6 oder 10 bis 12, **dadurch gekennzeichnet, dass** das Signal (330) der Direktschallquelle (813) aus dem Verhältnis der Summe beider Zeit-Frequenz-Darstellungen (115) und (125) zur Summe beider Panning-Koeffizienten (310) und (320) ermittelt wird.
14. Verfahren nach einem der Ansprüche 6 oder 10 bis 12, **dadurch gekennzeichnet, dass** die Umgebungssignale aus dem Verhältnis einer Differenz zwischen der Zeit-Frequenz-Darstellung (115) des ersten Audiosignals (110), gewichtet mit dem zweiten Panning-Koeffizienten (320), und der Zeit-Frequenz-Darstellung (125) des zweiten Audiosignals (120), gewichtet mit dem ersten Panning-Koeffizienten (310), zur Summe beider Panning-Koeffizienten (310) und (320) ermittelt wird.

Claims

1. A method for analyzing a stereo audio signal, wherein the stereo audio signal comprises a first audio signal (110) for a left-hand playback device (810) and a second audio signal (120) for a right-hand playback device (820), with the following steps:

the first audio signal (110) is converted into a first time-frequency representation (115), and the second audio signal (120) is converted into a second time-frequency representation (125);

from the first time-frequency representation (115), the time- and frequency-dependent power (115a) of the first audio signal (110) is determined (145), and from the second time-frequency representation (125), the time- and frequency-dependent power (125a) of the second audio signal (120) is determined (145);

a first equation is set up to relate the time- and frequency-dependent power (115a) of the first audio signal (110) to the product of the square of a first time- and frequency-dependent panning coefficient (310) with the time- and frequency-dependent power (330a) of a direct sound source (813) located in a listening area (890) between the left-hand playback device (810) and the right-hand playback device (820);

a second equation is set up to relate the time- and frequency-dependent power (125a) of the second audio signal (120) to the product of the square of a second time- and frequency-dependent panning coefficient (320) with the same time- and frequency-dependent power (330a) of the same direct sound source (813);

wherein the panning coefficients (310) and (320) are configured to position the direct sound source (813) within the listening area (890);

a position coefficient (390) that corresponds to the quotient of the difference of the panning coefficients (310)

and (320) divided by the sum of the panning coefficients (310) and (320) is determined as a solution of the equation system formed from the two equations,

characterized in that the equation system is solved under the further condition that the sum of the squares of the panning coefficients (310) and (320) is constant,

wherein the position coefficient (390) is determined from the quotient of the difference of the roots of the time- and frequency-dependent powers (125a, 115a) of both audio signals (110, 120) divided by the sum of the roots of the time- and frequency-dependent powers (125a, 115a) of both audio signals (110, 120).

2. The method of claim 1, **characterized in that** the first panning coefficient (310) is determined as the root of the quotient of the time- and frequency-dependent power (115a) of the first audio signal (110) divided by the sum of the time- and frequency-dependent powers (115a, 125a) of both audio signals (110, 120), and that the second panning coefficient (320) is determined as the root of the quotient of the time- and frequency-dependent power (125a) of the second audio signal (120) divided by the sum of the time- and frequency-dependent powers (115a, 125a) of both audio signals (110, 120).

3. The method of any one of claims 1 to 2, **characterized in that** the time- and frequency-dependent power (115a, 125a) of at least one audio signal (110, 120) at a point in time of interest is determined (145) as a weighted sum out of the time- and frequency-dependent power (115a, 125a) of the audio signal at an earlier point in time and out of the square of the time-frequency representation (115, 125) of this audio signal (110, 120) at the point in time of interest.

4. A method for analyzing a stereo audio signal, wherein the stereo audio signal comprises a first audio signal (110) for a left-hand playback device (810) and a second audio signal (120) for a right-hand playback device (820), with the following steps:

the first audio signal (110) is converted into a first time-frequency representation (115), and the second audio signal (120) is converted into a second time-frequency representation (125);

a first equation is set up to relate the first time-frequency representation (115) to the product of a first time- and frequency-dependent panning coefficient (310) with the time- and frequency-dependent signal (330) of a direct sound source (813) located in a listening area (890) between the left-hand playback device (810) and the right-hand playback device (820);

a second equation is set up to relate the second time-frequency representation (125) to the product of a second time- and frequency-dependent panning coefficient (320) with the same signal (330) of the same direct sound source (813);

wherein the panning coefficients (310) and (320) are configured to position the direct sound source (813) within the listening area (890);

a position coefficient (390) that corresponds to the difference of the squares of the panning coefficients (310) and (320) is determined as a solution of the equation system formed from the two equations, wherein the equation system is solved under the further condition that the sum of the squares of the panning coefficients (310 and (320) is constant,

characterized in that

the position coefficient (390) is determined from the quotient of the difference of the squared absolute values of both time-frequency representations (115) and (125) divided by the sum of the squared absolute values of both time-frequency representations (115) and (125).

5. The method of claim 4, **characterized in that** the first panning coefficient (310) is determined as the root of the quotient of the square of the time-frequency representation (115) of the first audio signal (110) divided by the sum of the squares of the time-frequency representations (115) and (125) of both audio signals (110) and (120), and that the second panning coefficient (320) is determined as the root of the quotient of the time-frequency representation (125) of the second audio signal (120) divided by the sum of the squares of the time-frequency representations (115) and (125) of both audio signals (110) and (120).

6. The method of any one of claims 1 to 5, **characterized in that** from the panning coefficients (310) and (320), the signal (330) of the direct sound source (813), and/or two ambient signals (510) and (520) that are not correlated with this direct sound source (813), are determined, wherein the first ambient signal (510) is only contained in the time-frequency representation (115) of the first audio signal (110), and the second ambient signal (520) is only contained in the time-frequency representation (125) of the second audio signal (120).

7. The method of claim 6, **characterized by** the following steps:

a first equation is set up to relate the first time-frequency representation (115) to the sum out of the product of the first panning coefficient (310) with the time- and frequency-dependent signal (330) of the direct sound source (813) and out of the filtering of a single common ambient signal (530) with a first decorrelation function (540);
 a second equation is set up to relate the second time-frequency representation (125) to the sum out of the product of the second panning coefficient (320) with the time- and frequency-dependent signal (330) of the direct sound source (813) and out of the filtering of the common ambient signal (530) with a second decorrelation function (550);
 the time- and frequency-dependent signal (330) of the direct sound source (813), and/or the common ambient signal, are determined as solutions of the equation system formed from both equations.

8. The method of claim 7, **characterized in that** the time- and frequency-dependent signal (330) of the direct sound source (813) is determined as a difference between the frequency band wise product of the first time-frequency representation (115) with the second decorrelation function (550) and the frequency band wise product of the second time-frequency representation (125) with the first decorrelation function (540), divided by the difference between the convolution of the first panning coefficient (310) with the second decorrelation function (550) and the frequency band wise product of the second panning coefficient (320) with the first decorrelation function (540).

9. The method of any one of claims 7 to 8, **characterized in that** the common ambient signal (530) is determined as a difference between the product of the second time-frequency representation (125) with the first panning coefficient (310) and the product of the first time-frequency representation (115) with the second panning coefficient (320), divided by the difference between the frequency band wise product of the first panning coefficient (310) with the second decorrelation function (550) and the frequency band wise product of the second panning coefficient (320) with the first decorrelation function (540).

10. The method of claim 6, **characterized in that** the signal (330) of the direct sound source (813) and the ambient signals (510, 520) are determined with an iterative method starting from an iteration rule that relates the signal (330) of the direct sound source of each iteration, and/or a contribution to this signal, to the ambient signals (510, 520) of the previous iteration.

11. The method of claim 10, **characterized in that** at each iteration, the panning coefficients (310) and (320) are recalculated from the ambient signals (510, 520) of the previous iteration.

12. The method of claim 11, **characterized in that** at each iteration, the first ambient signal (510) is corrected by an amount that is the product of the newly calculated first panning coefficient (310) with the signal (330) of the direct sound source (813) according to the current iteration, and that at each iteration, the second ambient signal (520) is corrected by an amount that is the product of the newly calculated second panning coefficient (320) with the signal (330) of the direct sound source (813) according to the current iteration.

13. The method of any one of claims 6 or 10 to 12, **characterized in that** the signal (330) of the direct sound source (813) is determined from the quotient of the sum of both time-frequency representations (115) and (125) divided by the sum of both panning coefficients (310) and (320).

14. The method of any one of claims 6 or 10 to 12, **characterized in that** the ambient signals are determined from the quotient of a difference between the time-frequency representation (115) of the first audio signal (110), weighted with the second panning coefficient (320), and the time-frequency representation (125) of the second audio signal (120), weighted with the first panning coefficient (310), divided by the sum of both panning coefficients (310) and (320).

Revendications

1. Procédé d'analyse d'un signal audio stéréo, le signal audio stéréo comportant un premier signal audio (110) destiné à un dispositif de reproduction gauche (810) et un deuxième signal audio (120) destiné à un dispositif de reproduction droit (820), le procédé comprenant les étapes suivantes :

le premier signal audio (110) est converti en une première représentation temps-fréquence (115) et le deuxième signal audio (120) est converti en une deuxième représentation temps-fréquence (125) ;

la puissance (115a), dépendant du temps et de la fréquence, du premier signal audio (110) est déterminée (145) à partir de la première représentation temps-fréquence (115) et la puissance (125a), dépendant du temps et de la fréquence, du deuxième signal audio (120) est déterminée (145) à partir de la deuxième représentation temps-fréquence (125) ;

une première équation est établie qui met en relation la puissance (115a), dépendant du temps et de la fréquence, du premier signal audio (110) avec le produit du carré d'un premier coefficient de panoramique (310), dépendant du temps et de la fréquence, et de la puissance (330a), dépendant du temps et de la fréquence, d'une source sonore directe (813) disposée dans une zone d'écoute (890) entre le dispositif de reproduction gauche (810) et le dispositif de reproduction droit (820) ;

une deuxième équation est établie qui met en relation la puissance (125a), dépendant du temps et de la fréquence, du deuxième signal audio (120) avec le produit du carré d'un deuxième coefficient de panoramique (320), dépendant du temps et de la fréquence, et de la même puissance (330a), dépendant du temps et de la fréquence, de la même source sonore directe (813) ;

les coefficients de panoramique (310) et (320) étant adaptés pour positionner la source sonore directe (813) dans la zone d'écoute (890) ;

un coefficient de position (390), qui correspond au rapport d'une différence entre les coefficients de panoramique (310) et (320) à la somme des coefficients de panoramique (310) et (320), étant déterminée comme solution du système d'équations formé à partir des deux équations,

caractérisé en ce que le système d'équations est résolu sous la condition supplémentaire que la somme des carrés des coefficients de panoramique (310) et (320) soit constante, le coefficient de position (390) étant déterminé à partir du rapport de la différence des racines des puissances (125a, 115a), dépendant du temps et de la fréquence, des deux signaux audio (110, 120) à la somme des racines des puissances (125a, 115a), dépendant du temps et de la fréquence, des deux signaux audio (110, 120).

2. Procédé selon la revendication 1, **caractérisé en ce que** le premier coefficient de panoramique (310) est déterminé comme étant la racine du rapport de la puissance (115a), dépendant du temps et de la fréquence, du premier signal audio (110) à la somme des puissances (115a, 125a), dépendant du temps et de la fréquence, des deux signaux audio (110, 120) et **en ce que** le deuxième coefficient de panoramique (320) est déterminé comme étant la racine du rapport de la puissance (125a), dépendant du temps et de la fréquence, du deuxième signal audio (120) à la somme des puissances (115a, 125a), dépendant du temps et de la fréquence, des deux signaux audio (110, 120).

3. Procédé selon l'une des revendications 1 à 2, **caractérisé en ce que** la puissance (115a, 125a), dépendant du temps et de la fréquence, d'au moins un signal audio (110, 120) à un instant intéressant est déterminée (145) comme étant la somme pondérée de la puissance (115a, 125a), dépendant du temps et de la fréquence, du signal audio à un instant antérieur et du carré de la représentation temps-fréquence (115, 125) de ce signal audio (110, 120) à l'instant intéressant.

4. Procédé d'analyse d'un signal audio stéréo, le signal audio stéréo comportant un premier signal audio (110) destiné à un dispositif de reproduction gauche (810) et un deuxième signal audio (120) destiné à un dispositif de reproduction droit (820), le procédé comprenant les étapes suivantes :

le premier signal audio (110) est converti en une première représentation temps-fréquence (115) et le deuxième signal audio (120) est converti en une deuxième représentation temps-fréquence (125) ;

une première équation est établie qui met en relation la première représentation temps-fréquence (115) avec le produit d'un premier coefficient de panoramique (310), dépendant du temps et de la fréquence, et du signal (330), dépendant du temps et de la fréquence, d'une source sonore directe (813) disposée dans une zone d'écoute (890) entre le dispositif de reproduction gauche (810) et le dispositif de reproduction droit (820) ;

une deuxième équation est établie qui met en relation la deuxième représentation temps-fréquence (125) avec le produit d'un deuxième coefficient de panoramique (320), dépendant du temps et de la fréquence, et du même signal (330) de la source sonore directe (813) ;

les coefficients de panoramique (310) et (320) étant adaptés pour positionner la source sonore directe (813) dans la zone d'écoute (890) ;

un coefficient de position (390), qui correspond à la différence entre les carrés des coefficients de panoramique (310) et (320), étant déterminé comme solution du système d'équations formé à partir des deux équations, le système d'équations étant résolu sous la condition supplémentaire que la somme des carrés des coefficients de panoramique (310) et (320) soit constante,

caractérisé en ce que le coefficient de position (390) est déterminé à partir du rapport de la différence des carrés des valeurs absolues des deux représentations temps-fréquence (115) et (125) à la somme des carrés des valeurs absolues des deux représentations temps-fréquence (115) et (125).

5 5. Procédé selon la revendication 4, **caractérisé en ce que** le premier coefficient de panoramique (310) est déterminé comme étant la racine du rapport du carré de la représentation temps-fréquence (115) du premier signal audio (110) à la somme des carrés des représentations temps-fréquence (115) et (125) des deux signaux audio (110) et (120) et **en ce que** le deuxième coefficient de panoramique (320) est déterminé comme étant la racine du rapport de la représentation temps-fréquence (125) du deuxième signal audio (120) à la somme des carrés des représentations temps-fréquence (115) et (125) des deux signaux audio (110) et (120).

10 6. Procédé selon l'une des revendications 1 à 5, **caractérisé en ce que** le signal (330) de la source sonore directe (813) et/ou deux signaux d'environnement (510) et (520) non corrélés à cette source sonore directe (813) sont déterminés à partir des coefficients de panoramique (310) et (320), le premier signal d'environnement (510) étant présent uniquement dans la représentation temps-fréquence (115) du premier signal audio (110) et le deuxième signal d'environnement (520) étant présent uniquement dans la représentation temps-fréquence (125) du deuxième signal audio (120).

15 7. Procédé selon la revendication 6, **caractérisé par** les étapes suivantes :

20 une première équation est établie qui met en relation la première représentation temps-fréquence (115) avec la somme du produit du premier coefficient de panoramique (310) avec le signal (330), dépendant du temps et de la fréquence, de la source sonore directe (813) et du filtrage d'un seul signal d'environnement commun (530) avec une première fonction de décorrélation (540) ;

25 une deuxième équation est établie qui met en relation la deuxième représentation temps-fréquence (125) avec la somme du produit du deuxième coefficient de panoramique (320) avec le signal (330), dépendant du temps et de la fréquence, de la source sonore directe (813) et du filtrage du signal d'environnement commun (530) avec une deuxième fonction de décorrélation (550) ;

30 le signal (330), dépendant du temps et de la fréquence, de la source sonore directe (813) et/ou le signal d'environnement commun (530) sont déterminés comme étant des solutions du système d'équations formé des deux équations.

35 8. Procédé selon la revendication 7, **caractérisé en ce que** le signal (330), dépendant du temps et de la fréquence, de la source sonore directe (813) est déterminé comme étant la différence entre le produit en bande de fréquences de la première représentation temps-fréquence (115) et de la deuxième fonction de décorrélation (550) et le produit en bande de fréquences de la deuxième représentation temps-fréquence (125) et de la première fonction de décorrélation (540), divisée par la différence entre la convolution du premier coefficient de panoramique (310) et de la deuxième fonction de décorrélation (550) et le produit en bande de fréquences du deuxième coefficient de panoramique (320) et de la première fonction de décorrélation (540).

40 9. Procédé selon l'une des revendications 7 à 8, **caractérisé en ce que** le signal d'environnement commun (530) est déterminé comme étant la différence entre le produit de la deuxième représentation temps-fréquence (125) et du premier coefficient de panoramique (310) et le produit de la première représentation temps-fréquence (115) et du deuxième coefficient de panoramique (320), divisée par la différence entre le produit en bande de fréquences du premier coefficient de panoramique (310) et de la deuxième fonction de décorrélation (550) et le produit en bande de fréquences du deuxième coefficient de panoramique (320) et de la première fonction de décorrélation (540).

45 10. Procédé selon la revendication 6, **caractérisé en ce que** le signal (330) de la source sonore directe (813) et les signaux d'environnement (510, 520) sont déterminés avec un procédé itératif à partir d'une règle d'itération qui met en relation le signal (330) de la source sonore directe de chaque itération et/ou une contribution à ce signal avec les signaux d'environnement (510, 520) de l'itération précédente.

50 11. Procédé selon la revendication 10, **caractérisé en ce que** les coefficients de panoramique (310) et (320) sont recalculés à chaque itération à partir des signaux d'environnement (510, 520) de l'itération précédente.

55 12. Procédé selon la revendication 11, **caractérisé en ce que** le premier signal d'environnement (510) est corrigé à chaque itération d'une valeur qui est le produit du premier coefficient de panoramique recalculé (310) et du signal (330) de la source sonore directe (813) en fonction de l'itération en cours, et **en ce que** le deuxième signal d'environnement (520) est corrigé à chaque itération d'une valeur qui est le produit du deuxième coefficient de panoramique recalculé (320) et du signal (330) de la source sonore directe (813) en fonction de l'itération en cours.

ronnement (520) est corrigé à chaque itération d'une valeur qui est le produit du deuxième coefficient de panoramique recalculé (320) et du signal (330) de la source sonore directe (813) en fonction de l'itération en cours.

5 13. Procédé selon l'une des revendications 6 ou 10 à 12, **caractérisé en ce que** le signal (330) de la source sonore directe (813) est déterminé à partir du rapport de la somme des deux représentations temps-fréquence (115) et (125) à la somme des deux coefficients panoramiques (310) et (320).

10 14. Procédé selon l'une des revendications 6 ou 10 à 12, **caractérisé en ce que** les signaux d'environnement sont déterminés à partir du rapport d'une différence entre la représentation temps-fréquence (115) du premier signal audio (110), pondérée avec le deuxième coefficient de panoramique (320), et de la représentation temps-fréquence (125) du deuxième signal audio (120), pondérée avec le premier coefficient de panoramique (310), à la somme des deux coefficients de panoramique (310) et (320).

15

20

25

30

35

40

45

50

55

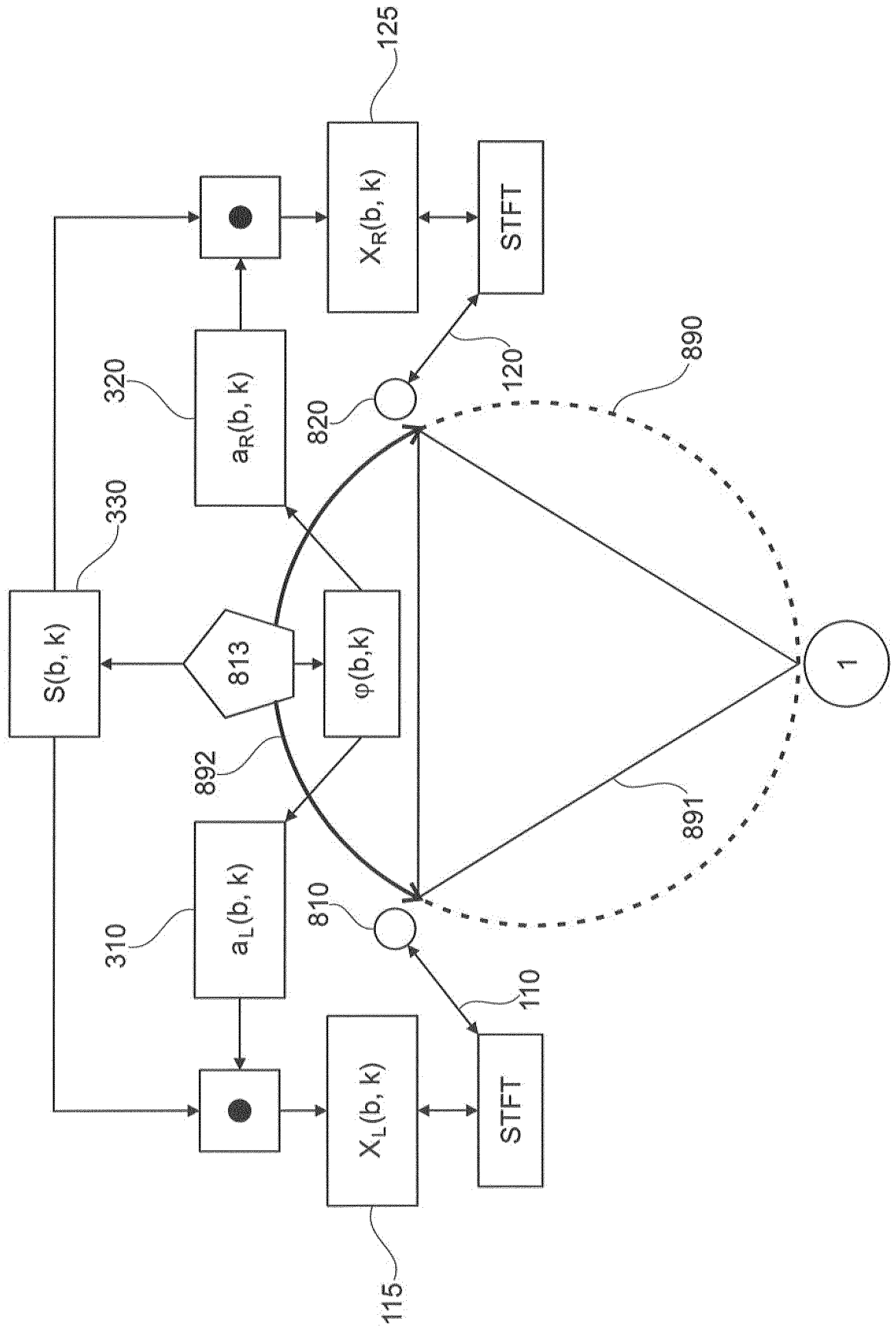


Fig. 1

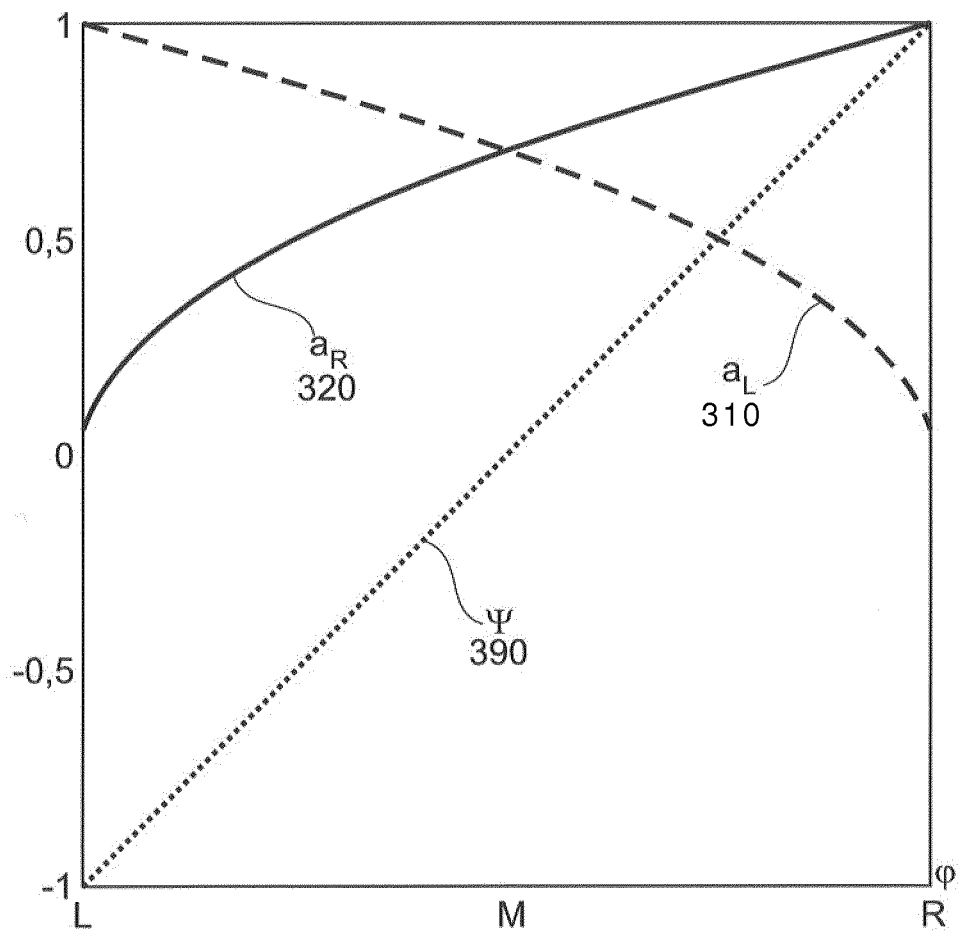


Fig. 2

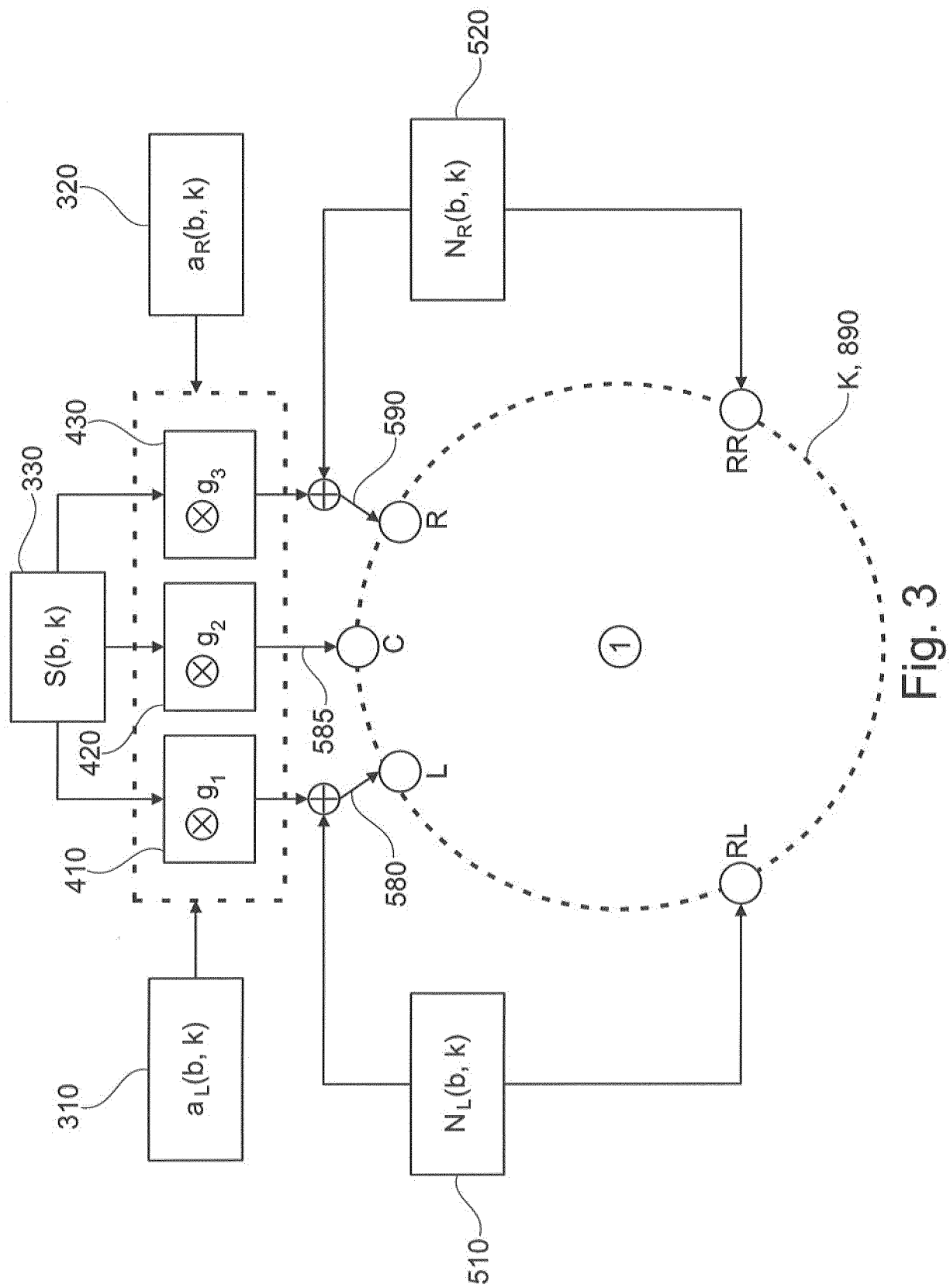


Fig. 3

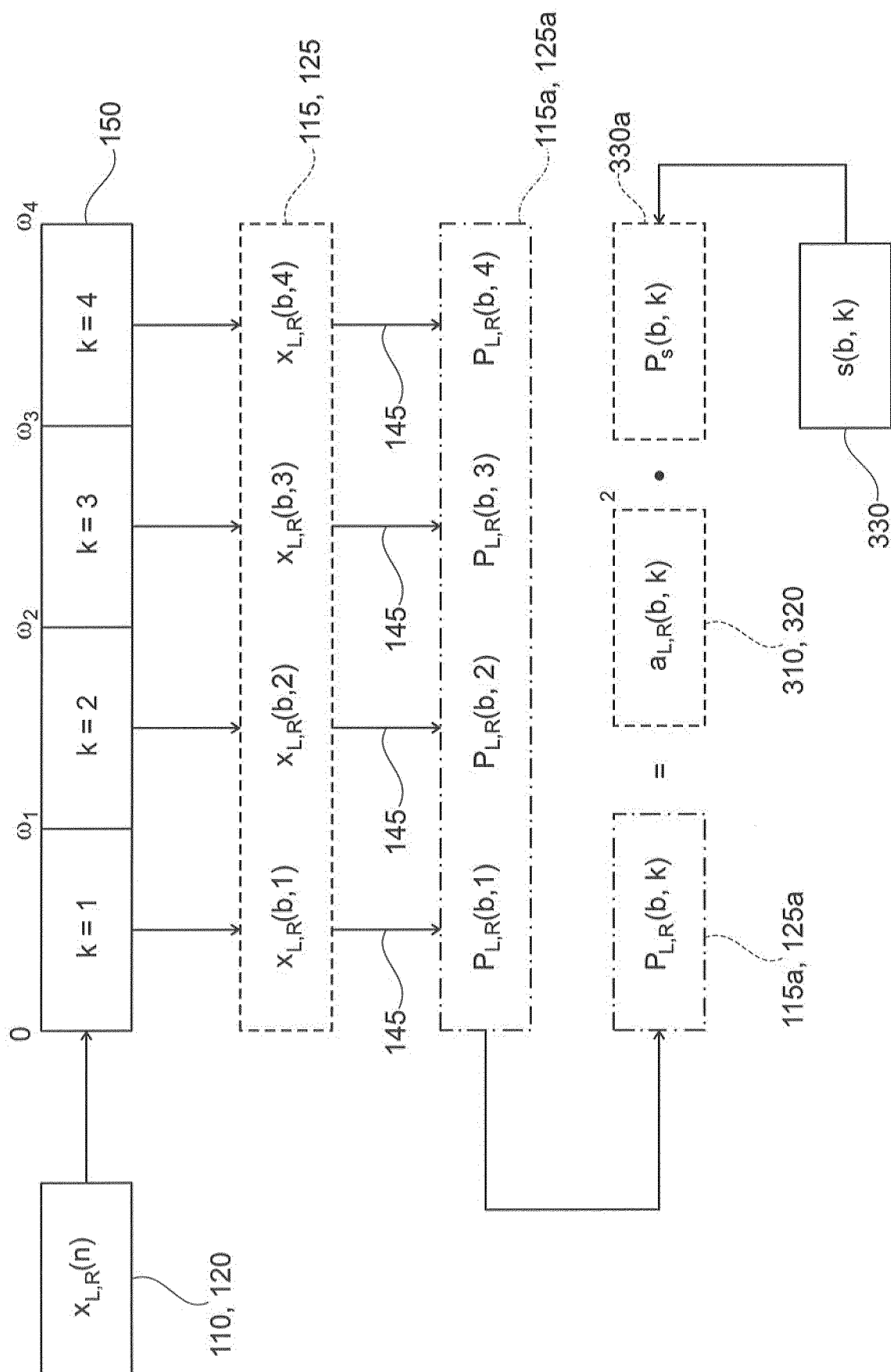


Fig. 4

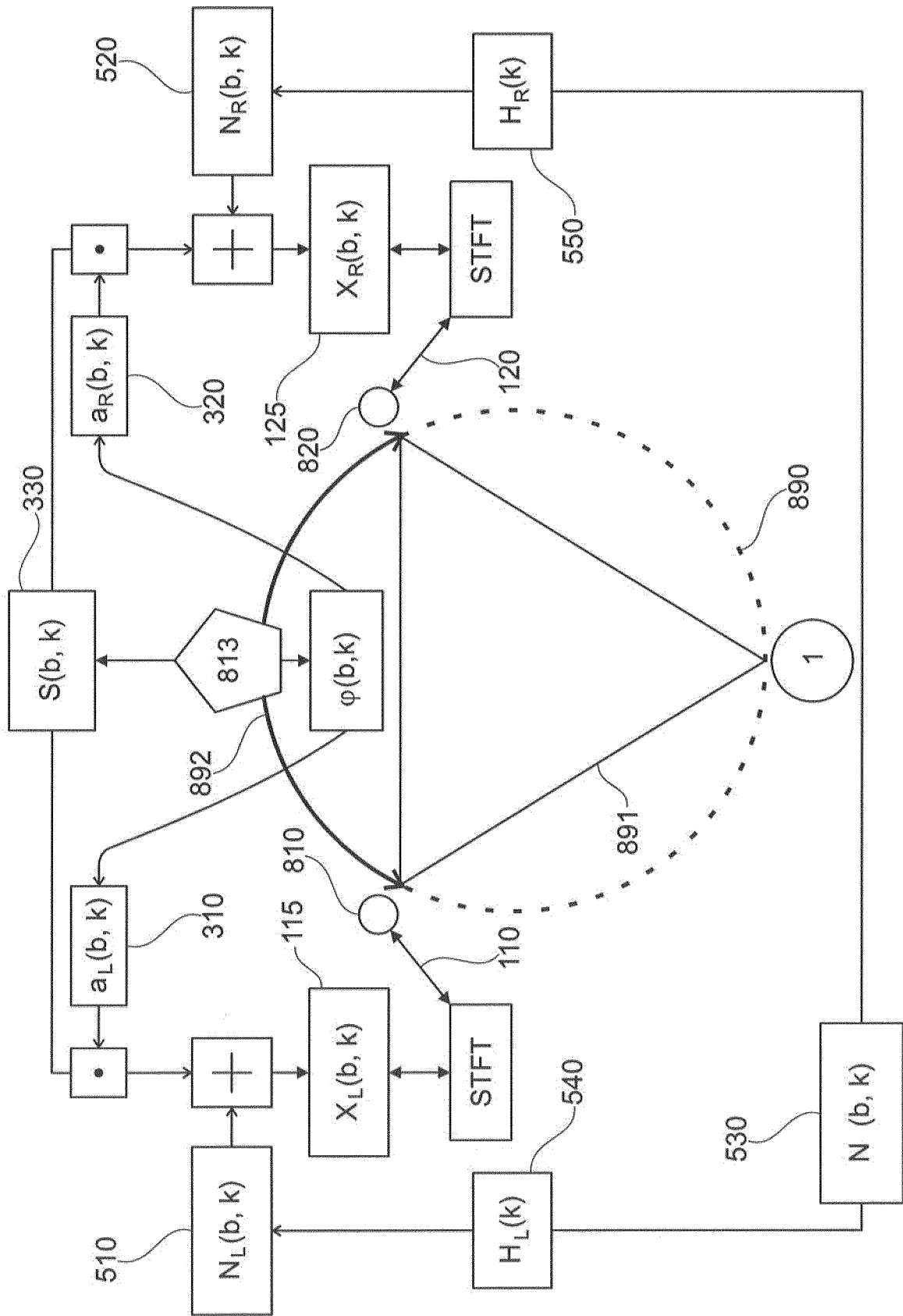


Fig. 5

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

- DE 102012017296 B4 [0005] [0051]
- WO 2010028784 A1 [0006]

In der Beschreibung aufgeführte Nicht-Patentliteratur

- "vector base amplitude panning" gemäß (Ville Pulkki, "Virtual sound source positioning using vector based amplitude panning. *Journal of the Audio Engineering Society*, Juni 1997, vol. 45 (6), 456-466 [0051]