



(11)

EP 2 357 854 B1

(12)

EUROPÄISCHE PATENTSCHRIFT

(45) Veröffentlichungstag und Bekanntmachung des
Hinweises auf die Patenterteilung:
09.03.2016 Patentblatt 2016/10

(51) Int Cl.:
H04S 7/00 (2006.01)

(21) Anmeldenummer: **10197378.2**

(22) Anmeldetag: **30.12.2010**

(54) **Verfahren und Vorrichtung zur Erzeugung individuell anpassbarer binauraler Audiosignale**

Method and device for generating individually adjustable binaural audio signals

Procédé et dispositif de production de signaux audio binauraux individuellement adaptables

(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**

(30) Priorität: **07.01.2010 DE 102010004171**

(43) Veröffentlichungstag der Anmeldung:
17.08.2011 Patentblatt 2011/33

(73) Patentinhaber: **Deutsche Telekom AG
53113 Bonn (DE)**

(72) Erfinder:
• **Lindau, Alexander
10437 Berlin (DE)**
• **Estrella Cazuriaga, Rolando Jorgos
10961 Berlin (DE)**

(74) Vertreter: **2K Patentanwälte Blasberg Kewitz &
Reichel
Partnerschaft mbB
Schumannstrasse 27
60325 Frankfurt am Main (DE)**

(56) Entgegenhaltungen:
**US-A1- 2006 045 294 US-B1- 6 795 556
US-B1- 7 174 229**

- **DEFRANCE GUILLAUME ET AL: "Finding the onset of a room impulse response: Straightforward?", THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, AMERICAN INSTITUTE OF PHYSICS FOR THE ACOUSTICAL SOCIETY OF AMERICA, NEW YORK, NY, US, Bd. 124, Nr. 4, 24. September 2008 (2008-09-24), Seiten 248-254, XP012120513, ISSN: 0001-4966, DOI: DOI:10.1121/1.2960935**

EP 2 357 854 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 und eine Vorrichtung zur Erzeugung individuell anpassbarer binauraler Audiosignale.

[0002] Die Realisierung plausibler oder gar authentischer virtueller akustischer Umgebungen ist seit längerer Zeit Ziel technischer Entwicklung im Bereich binauraler Audiosysteme und -verfahren. Diese Verfahren zielen auf die Erzeugung eines möglichst naturgetreuen räumlichen Höreindrucks beim Hörer (Nutzer, Versuchsperson, Patient usw.). Die wesentlichen Parameter bzw. sog. cues (Hinweise) für räumliches Hörens stellen -richtungs- und frequenzabhängige Laufzeit- und Pegeldifferenzen dar, die die Schallsignale einer Quelle beim Eintreffen an beiden Ohren bzw. den Trommelfellen eines Hörers aufweisen. Hierzu sei verwiesen auf die sog. Duplextheorie des Hörens nach Lord Rayleigh.

[0003] Diese Parameter müssen von einem System zur Simulation virtueller (Raum-)Akustik (auch kurz VAE bzw. virtual acoustic environment genannt) korrekt hergestellt werden können. Im Allgemeinen spricht man dabei von der interauralen Laufzeitdifferenz (auch kurz ITD bzw. interaural time delay genannt) oder von der interauralen Pegeldifferenz (auch kurz ILD bzw. interaural level difference genannt). Letztere ist im Wesentlichen auf den Kopf als Hindernis bei der Schallausbreitung zurückzuführen, feinere Unterschiede lassen sich auf die spezifische Form der Außenohren zurückführen. Die interaurale Laufzeitdifferenz ITD ergibt sich dagegen hauptsächlich aus dem geometrischen Abstand beider Ohren, welcher positionsabhängig zu unterschiedlichen Eintreffzeitpunkten des Schalls einer Quelle an den Trommelfellen führt. Die ITD lässt sich nur innerhalb eines Frequenzbereichs bis ca. 1500 Hz sinnvoll für die Lokalisation auswerten, oberhalb davon verhindern Ambiguitäten diese Auswertung und die Pegeldifferenz ILD bestimmt zunehmend die wahrgenommene Schalleinfallrichtung. Beide Aspekte sind teilweise gegeneinander austauschbar (sog. "Laufzeit-Intensitäts-Trading"), wovon z.B. die Stereophonie profitiert, die so wahlweise als Laufzeit-, Pegel- oder gemischte Stereophonie realisierbar ist.

[0004] Eine vollständige, frequenzabhängige Beschreibung sowohl der zeitlichen als auch der spektralen Informationen des Schallfelds stellt die kopfbezogene oder Außenohr-Übertragungsfunktion dar. Sie ist definiert als das Verhältnis der folgenden Übertragungsfunktionen: Zum einen der Übertragungsfunktion gemessen am Trommelfell eines Hörers (wahlweise auch am geblockten Ohrkanaleingang eines Kunstkopfes) und zum zweiten der Übertragungsfunktion eines Kugelmikrofons in Kopfmittle bei dessen Abwesenheit. Diese sogenannte kopfbezogene Übertragungsfunktion (auch kurz HRIR bzw. HRTF genannt, wobei HRIR für head related impulse response und HRTF für head related transfer function stehen) ist per Definition nachhallfrei. In der GB-A-2 369 976 wird beispielsweise ein Verfahren zur Synthese einer gemittelten Diffus-Feld-HRTF beschrieben. Ein Verfahren und eine Vorrichtung zur Verarbeitung von interauraler Zeitverschiebung ("Method and Apparatus for Processing Interaural Time Delay in 3D Digital Audio") wird in US 7,174,229 B1 beschrieben. In US 6,795,556 wird die Modifizierung von kopfbezogenen Übertragungsfunktion (HRTFs) beschrieben. Verfahren zur Messung von HRTFs werden in der EP 0912077 B1 oder der DE 69523643 T2 beschrieben.

[0005] Werden hingegen Kunstkopfübertragungsfunktionen einer elektro-akustischen Quelle in einem nachhallbehafteten Raumschallfeld vermessen, so spricht man unterscheidend dazu von BRIRs oder BRTFs (binaural room impulse responses bzw. transfer functions), also von kopfbezogenen Übertragungsfunktionen, die auch den nachhallenden Raum mit umfassen. Mittels einer solchen kopfbezogenen Übertragungsfunktion kann man den akustischen Eindruck sog. virtueller Schallquellen mit hoher Plausibilität erzeugen. Filtert man nämlich ein beliebiges nachhallfreies Audiosignal mit den HRTF des linken und rechten Ohres, stellt sich bei der Wiedergabe z.B. über Kopfhörer ein Schallereignis mit mehr oder minder korrektem Richtungs- und Entfernungseindruck ein. Etwas vereinfachend besagt die Theorie der Binauraltechnik, dass die auf diese Weise mögliche, perfekte Rekonstruktion des Schalldruckzeitverlaufs am Trommelfell zu einem, dem realen Hören eines identischen Hörereignisses führt.

[0006] An sich bekannt sind rechnerbezogene Verfahren und Systeme für numerische Raumakustiksimulation, die Datenbanken mit HRTF-Datensätzen eines individuell vermessenen Hörers oder eines Kunstkopfes in hoher räumlicher Winkelauflösung aufweisen und die das Berechnen der vollständigen Raumimpulsantwort an ausgewählten Hörerplätzen ermöglichen, jedoch idealerweise nur für diesen individuellen Hörer oder für einen Hörer, dessen Physiognomie bzw. Anthropometrie nur unwesentlich von der des verwendeten Kunstkopfes abweicht. Dazu werden, vereinfacht gesagt, unter Verwendung eines Strahlenmodells aus der Optik, mögliche Schallpfade von der Quelle zum Hörer nachgebildet. Jeder Schallpfad entspricht einer Raumreflexion; durch Gewichtung der Schallpfade mit der der Einfallrichtung des Schallpfades entsprechenden HRTF des linken und rechten Ohres und nach Superpositionierung aller derart gewichteten Reflexionen der Raumimpulsantwort ergibt sich eine binaurale Raumimpulsantwort des virtuellen Raums. Dieser Raum kann nun durch Filterung nachhallfreier Audiomaterials mit dieser BRIR hörbar gemacht werden; man spricht dann auch von "Auralisation".

[0007] Alternativ dazu können auch die BRIRs direkt in-situ gemessen werden. Dazu werden die BRIRs unter Verwendung elektroakustischer Anregequellen (Lautsprecher) und eines sog. Kunstkopf-Torsosimulators bzw. Roboters (auch kurz als HATS bezeichnet; head and torso simulator) vermessen. Ein solcher Roboter (HATS) ermöglicht aufgrund einer Nachbildung der multiplen Freiheitsgrade der Kopfbewegung (z.B. 4 rotatorische, 3 translatorische) eine automatisierte und räumlich feine Vermessung der BRIRs für Kopfbewegungen eines Hörers und kann eine deutlich höhere

Klangqualität und Nähe zum Original erzielen. In dem Artikel Lindau, A.; Weinzierl, S. (2007): "Fabian - Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden." von A. Lindau und S. Weinzierl, erschienen in "Fort-schritte der Akustik", Tagungsband der 33. DAGA Stuttgart, Deutschland, wird eine solche Technologie unter der Bez-
 5 eichnung FABIAN vorgestellt. Der HATS FABIAN wird durch einen Softwaregesteuerten automatisierbaren Roboter
 realisiert, der aus einem mit Außenohren abgeformtem Kunstkopf und generisch nachempfundenem menschlichem
 Körper besteht. Der Roboter dient zum Erreichen natürlicher Schallfeldbeeinflussung (Beugung, Abschattung, Reflektion)
 als Resultat am Ort der eigentlichen Messung des Schallfelds, der beiden Mikrofonmembranen am geblockten Ohrkanal.
 Durch Verwendung eines solchen HATS kann eine verbesserte Auralisation erreicht werden.

[0008] Eine weitere Verbesserung durch Erhöhung der Plausibilität binauraler raumakustischer Simulation ergibt sich,
 10 wenn die Interaktivität des Hörers berücksichtigt wird, d.h. wenn das Reagieren der Simulation auf Kopfbewegungen
 des Hörers berücksichtigt wird. Wünschenswert wäre es, wenn jedwede gewollte oder unbewusste Kopfbewegung kom-
 pensiert werden könnte, um somit zu einem plausiblen und fehlerfreien räumlichen Hören beizutragen. Dazu müssten
 jedoch für jede in einer späteren Auralisation zu berücksichtigenden Kopfstellung des Hörers (evtl. bzgl. verschiedener
 15 translatorischer und rotatorischer Freiheitsgrade) die kopfbezogenen Übertragungsfunktionen als HRTFs bzw. BRIRs
 vorliegen. Bekannt ist bereits, eine auf moderner Computerhardware ausgeführte Implementation echtzeitfähiger, zeit-
 varianter schneller Faltungsalgorithmen, die den Kopfbewegungen des Hörers (durch sog. head tracking) folgend, die
 HRTFs oder BRIRs unhörbar schnell austauschen, so dass sich z.B. im Falle von HRTFs relative Bewegungen zwischen
 Quelle und Empfänger oder im Falle von BRIRs ein Raumschallfeld, welches sich bei Kopfbewegungen nicht mehr
 "mitdreht" simulieren lassen. Zudem können auch typische Abbildungsartefakte der sog. statischen Kunstkopftechnik,
 20 wie "Vorne-Hinten-Vertauschungen" und "Im-Kopf-Lokalisation", wirkungsvoll eliminiert bzw. reduziert werden.

[0009] In der US 2006/045294 A1 wird ein Verfahren zur Erzeugung individuell anpassbarer Audiosignale beschrieben,
 bei dem ein Hörer (listener) den Klang von virtuellen Lautsprechern über Kopfhörer mit einem realitätsnahen Höreindruck
 erleben kann, der schwer von dem Höreindruck echter Lautsprecher zu unterscheiden ist. Dazu werden mehrere per-
 sonalisierte Raumimpulsantworten (PRIRs) für die Lautsprecher-Schallquellen über eine begrenzte Anzahl von Zuhörer-
 25 Kopfpositionen ermittelt. Dann werden die PRIRs verwendet, um ein Audiosignal für die Lautsprecher in eine virtualisierte
 Ausgabe für den Kopfhörer zu wandeln bzw. zu transformieren. Indem die Transformation auf der Kopfposition des
 Hörers (listeners) basiert, kann das System die Transformation so anpassen, dass die virtuellen Lautsprecher sich nicht
 zu bewegen scheinen, wenn der Hörer seinen Kopf bewegt.

[0010] Der Artikel "Finding the onset of a room impulse response: Straightforward?" von G. Defrance et al., erschienen
 30 in "The Journal of the Acoustical Society of America (JASA)", Band 124(4), Seiten 248-254, Ausgabe Oct. 2008, Online
 publiziert am 24. Sep. 2008, befasst sich mit der genauen Bestimmung des Onset (Beginn der Audiosignale bzw. -
 daten) in Bezug auf die zeitliche Dimension (timing) der räumlichen Impulsantworten (room impulse responses: RIRs).

[0011] In der US 7174229 B1 werden ein Verfahren und eine Vorrichtung zur Verarbeitung interauraler Zeitverzögerung
 35 in 3D digitalen Audiosignalen bzw. -daten beschrieben. Für das Wiedereinbringen extrahierter oder modellierter ITDs
 in HRTF-Datensätzen (keine Räume) wird allerdings ein umständliches zweistufiges Delay aus samplegenauen FIFO-
 Puffern und fraktionalen Allpassfiltern (Auftrennung von sample- und subsample-delay) verwendet. Somit benötigt dieser
 Vorschlag eine ungenannte Anzahl von explizit für jede fraktionale Zeiteinheit entworfener Filter, die z.B. durch Paral-
 lelschaltung realisiert werden und durch Anwahl mit einem Schalter selektierbar gemacht werden müssen.

[0012] In der US 6795556 wird ein Verfahren beschrieben, das auf eine Skalierung von empirischen "non standard"
 40 HRTF-Cues abzielt, sowohl im Zeit- als auch im Frequenzbereich ab. Auch dieses Verfahren geht nicht über das zuvor
 beschriebene hinaus.

[0013] Obwohl die bekannten Verfahren und Vorrichtungen bereits mittels datenbasierter, dynamischer binauraler
 Simulation eine hochplausible Simulation von Raumakustik ermöglichen, bestehen noch viele Herausforderungen, um
 zu einer perzeptiven Perfektionierung gelangen zu können. Hier sind insbesondere folgende Aufgabenbereiche zu
 45 nennen:

Die Wahrnehmbarkeit von Latenz bei der Verarbeitung der Kopfbewegung, die minimal notwendige Granularität
 des Gridrasters der BRIRs bzgl. Kopfbewegungen in allen rotatorischen Freiheitsgraden, die biskriminierbarkeit
 50 räumlich getrennter Quellen, beschleunigte BRIR-Messverfahren, die Simulation natürlicher Quellrichtcharakteris-
 tika (Sprecher, Instrumente) und/oder die optimale Frequenzgangsentzerrung der Wiedergabestrecke.

[0014] Im Kreise der Erfinder wurde eine Evaluation der Simulationsumgebung durch direkten A-B-Vergleich (also
 Vergleich unter direktem Schalten von einer Quelle zu einer anderen, um einen unbeeinflussten und möglichst objektiven.
 Vergleich zu ermöglichen) von simuliertem und realem Schallfeld durchgeführt, die neben einer erfreulich geringen
 55 Erkennungsrate von 52.6% leider auch eine Anzahl verbleibender perzeptiver Artefakte zeigte, von denen als besonderes
 prägnante Probleme das Auftreten von Klangverfärbung und Lokalisationsfehler hervortreten.

[0015] Demnach ist es Aufgabe der Erfindung, ein Verfahren der eingangs genannten Art vorzuschlagen, das die
 zuvor genannten Probleme in vorteilhafter Weise überwindet. Insbesondere soll das Auftreten perzeptiver Artefakte wie

beispielsweise von Lokalisationsfehlern, Latenzwahrnehmung oder Überblendungsfehlern eliminiert oder zumindest deutlich reduziert werden. Außerdem soll eine Vorrichtung zur Durchführung des Verfahrens vorgeschlagen werden.

[0016] Gelöst wird die Aufgabe durch ein Verfahren mit den Merkmalen des Anspruchs 1 sowie durch eine Vorrichtung mit den Merkmalen des nebengeordneten Anspruchs.

[0017] Demnach wird ein Verfahren vorgeschlagen, bei dem zur Erzeugung individuell anpassbarer binauraler Audiosignale folgende Schritte ausgeführt werden:

aus kopfbezogenen Übertragungsfunktionen, die verschiedenen Kopfpositionen zugeordnet sind, werden mittels Extraktion von interauralen Laufzeiten laufzeitbefreite kopfbezogene Übertragungsfunktionen und dazu korrespondierende Laufzeit-Differenzen berechnet;

mittels einer Binauralsynthese werden laufzeitbefreite Audiosignale erzeugt, die sich jeweils auf einen binauralen Signalpfad beziehen, wobei in Abhängigkeit von Kopfpositions-Daten die laufzeitbefreiten kopfbezogenen Übertragungsfunktionen für eine aktuelle Kopfposition verwendet werden; und

aus den laufzeitbefreiten Audiosignalen werden die individuell anpassbaren binauralen Audiosignale erzeugt, indem zumindest für einen Signalpfad das erzeugte laufzeitbefreite Audiosignal mit einer zeitlichen Verzögerung in Abhängigkeit von der für die aktuelle Kopfposition berechneten Laufzeit-Differenz und in Abhängigkeit eines individuellen Skalierungsfaktors beaufschlagt wird,

wobei bei einer auftretenden Änderung von einer vorherigen Kopfposition in die aktuelle Kopfposition die zeitliche Verzögerung zwischen einem ersten Wert und einem zweiten Wert adaptiv mittels einer Abtastratenkonversion verändert wird, wobei die Abtastratenkonversion einen Konversionsfaktor verwendet um das laufzeitbefreite Audiosignal um den Konversionsfaktor beschleunigt oder verzögert abzuspielen, und wobei der für die Abtastratenkonversion verwendete Konversionsfaktor entsprechend der Änderung von der zu der vorherigen und zu der aktuellen Kopfposition gehörenden zeitlichen Verzögerung ermittelt wird.

[0018] Ebenfalls wird eine Vorrichtung zur Durchführung des Verfahrens vorgeschlagen, wobei die Vorrichtung zur Erzeugung räumlich wahrnehmbarer binauraler Audiosignale folgende Komponenten aufweist:

eine erste Schaltung, die aus kopfbezogenen Übertragungsfunktionen, die verschiedenen Kopfpositionen zugeordnet sind, interaurale Laufzeiten extrahiert und daraus Laufzeit-Differenzen berechnet;

eine zweite Schaltung, die mittels einer Binauralsynthese laufzeitbefreite Audiosignale erzeugt, die sich jeweils auf einen binauralen Signalpfad beziehen, wobei diese Schaltung in Abhängigkeit von Kopfpositions-Daten die laufzeitbefreiten kopfbezogenen Übertragungsfunktionen für eine aktuelle Kopfposition verwendet; und

mindestens eine dritte Schaltung, die aus den laufzeitbefreiten Audiosignalen die individuell anpassbaren binauralen Audiosignale erzeugt, indem diese Schaltung zumindest für einen Signalpfad das erzeugte laufzeitbefreite Audiosignal mit einer zeitlichen Verzögerung in Abhängigkeit von der für die aktuelle Kopfposition berechneten Laufzeit-Differenz und in Abhängigkeit eines individuellen Skalierungsfaktors beaufschlagt wobei sie in der Lage ist, die kopfpositionsbedingten Veränderungen der Laufzeitdifferenzen im Signal auf unhörbare Weise zu vollziehen.

[0019] Demnach wird vorgeschlagen, aus kopfbezogenen Übertragungsfunktionen, wie z.B. BRIRs, die interauralen Laufzeiten zu extrahieren, um somit laufzeitbefreite Übertragungsfunktionen zu erhalten, und aus den extrahierten Laufzeiten Laufzeitdifferenzen zu berechnen, die in einem späteren Schritt zusammen mit einem individuellen Skalierungsfaktor verwendet werden, um die Audiosignale damit zu beaufschlagen, welche durch eine Binauralsynthese in Abhängigkeit der aktuellen Kopfposition mittels der laufzeitbefreiten Übertragungsfunktionen erzeugt wurden. Demzufolge wird in einem mehrstufigen Verfahren die kopfbezogene Übertragungsfunktionen von den interauralen Laufzeiten befreit, die sich auf eine bestimmte Anthropometrie (z.B. die eines Kunstkopfes) beziehen, und es werden laufzeitbefreite Audiosignale mittels einer Binauralsynthese erzeugt und anschließend durch eine zeitliche Verzögerung, die individuell gewichtet wird (Skalierungsfaktor), mit den für die jeweilige Person bzw. den Nutzer passenden Laufzeiten beaufschlagt, so dass ein deutlich verbesserter räumlicher Höreindruck erzielt werden kann. Insbesondere entspricht die zeitliche Verzögerung einem Wert, der aus der Wichtung der für die aktuelle Kopfposition berechneten Laufzeit-Differenz und dem individuellen Skalierungsfaktor berechnet wird.

[0020] Das Verfahren wird dynamisch gestaltet, indem bei einer auftretenden Änderung von einer vorherigen Kopfposition in die aktuelle Kopfposition die zeitliche Verzögerung zwischen einem ersten Wert und einem zweiten Wert adaptiv mittels einer Abtastratenkonversion bzw. sample rate conversion (SRC) verändert wird. Dabei verwendet die Abtastratenkonversion einen Konversionsfaktor um das laufzeitbefreite Audiosignal um den Konversionsfaktor beschleunigt oder verzögert abzuspielen, wobei der für die Abtastratenkonversion verwendete Konversionsfaktor entsprechend der Änderung von der zu der vorherigen und zu der aktuellen Kopfposition gehörenden zeitlichen Verzögerung ermittelt wird.

[0021] Durch die Anwendung eines SRC-Verfahrens, das auf Basis zeitlich kurzer Audioblöcke - wahrgenommen

latenzlos - arbeitet, ist eine nahtlose Änderung der Laufzeitdifferenz bei Kopfbewegungen möglich.

[0022] Mit dem Verfahren können die laufzeitbefreiten kopfbezogenen Übertragungsfunktionen für eine Vielzahl von Kopfpositionen und/oder für eine Vielzahl von Audiosignal-Quellen berechnet werden. Bei der Vielzahl von Kopfpositionen kann es sich z.B. um die Auflösung der Blickrichtung in kleinen Winkeländerungen bzw. -schritten, von z.B. 1 Grad, handeln. Die Anzahl der Quellen (Schall- bzw. Audioquellen) kann sehr groß sein, wobei pro Quelle und für jeden Signalpfad (linker und rechter Ohrkanal) ein Audiosignal erzeugt wird. Diese können dann nach der individuellen Gewichtung (Laufzeit-Verzögerung) für jeden Signalpfad superpositioniert werden.

[0023] Die zu Beginn des Verfahren durchgeführte Extraktion der interauralen Laufzeiten aus den kopfbezogenen Übertragungsfunktionen kann z.B. mittels eines der folgenden Verfahren durchgeführt werden: Onset-Verfahren, interaurales Kreuzkorrelationsverfahren, frequenzabhängige Gruppenlaufzeitdifferenzbildung, Differenzbildung der frequenzabhängig linear angenäherten Phasengradienten oder Bestimmung des Exzessphasenanteils aus Division der Ü.-funktionen vor und nach Hilberttransformation. Insbesondere das Onset-Verfahren führt zu sehr guten Ergebnissen, was später noch im Detail beschrieben wird.

[0024] Die Erfindung löst in vorteilhafter Weise das Problem, dass bei Auralisationsverfahren normalerweise die kopfbezogenen Übertragungsfunktionen HRTFs oder BRIRs immer nur für die Anthropometrie eines bestimmten Individuums bzw. für einen bestimmten Kunstkopf gelten, wodurch individuellen Unterschiede, wie z.B. die des Kopfdurchmessers nicht exakt durch die in den Übertragungsfunktionen enthaltenen Laufzeitinformationen repräsentiert werden, was dazu führt, dass andere Hörer, also "fremde" Personen, eine mehr oder minder stark verzerrte Wahrnehmung von Lokalisation und - bei Kopfbewegungen - die Empfindung einer naturgemäß nicht vorhandenen, räumlichen Bewegung der Audiosignale (Lokalisationsinstabilität) erfahren.

[0025] Durch die Erfindung werden also Lokalisationsfehler aufgrund eines falschen Kopfdurchmessers (Abweichung vom Kunstkopf) vermieden. Als Nebenaspekt der gewählten Vorgehensweise (Laufzeitbefreiung und quasi-minimalphasige Überblendung) ergeben sich zudem deutliche Verbesserungen hinsichtlich der Latenzaspekte und der hörbaren Fehler bei der Überblendung bei Kopfbewegungen.

[0026] Untersuchungen der Erfinder haben gezeigt, dass die auftretenden Artefakte unterschiedlich starken Einfluss auf die Authentizität der Hörwahrnehmung haben. So können offensichtliche Klangfarbenfehler ohne direkte Referenz schnell unmerkbar werden, da beim Hörer eine Gewöhnung eintritt; das absolute Klangfarbengedächtnis ist im Allgemeinen eher schlecht. Schwerwiegender erscheint die Adaption an Abweichungen in den Kopfdimensionen und die daraus resultierenden falschen Laufzeitinformationen von Schallsignalen. So führt ein zu großer Kopfdurchmesser im generischen Datensatz zu einer wahrgenommen Bewegung frontaler Quellen in entgegengesetzte Kopfbewegungsrichtung; im umgekehrten Falle "folgen" die Quellen scheinbar in geringem Maße den Bewegungen des Hörers. Dieses Artefakt kann sehr störend sein und Adaptionfähigkeiten sind bisher nicht untersucht. Die Erfindung reduziert auch die normalerweise auftretenden Überblendungsfehler ("Stottern"), die dadurch entstehen, dass bei einem Echtzeitaustausch der HRTFs bzw. BRIRs Filter mit Laufzeitversätzen ineinander überblendet werden. Dieses zeitliche "misalignment" führt zu typischen kammfilterartigen Überblendartefakten, die besonders bei quasistationären Inhalten (bei Sprachanwendungen z.B. bei Vokalen, bei Musik z.B. bei "Streicherteppichen") deutlich und störend zu Tage treten. Die Erfindung reduziert diese Überblendungsfehler, indem die Überblendung der Übertragungsfunktionen und das Einfügen von Laufzeitdifferenzen bei der Binauralsynthese zeitlich aufeinander folgend und nicht - wie herkömmlich üblich - zeitgleich erfolgen.

[0027] Nachfolgend wird die Erfindung und die sich daraus ergebenden Vorteile anhand von Ausführungsbeispielen und mit Bezug auf die beiliegenden Zeichnungen beschrieben, die folgende schematische Darstellungen wiedergeben:

Fig. 1 zeigt das Ablaufdiagramm für ein erfindungsgemäßes Verfahren;

Fig. 2 veranschaulicht die Extraktion von interauralen Laufzeiten und die Berechnung von individuellen Laufzeitdifferenzen;

Fig. 3 zeigt als Blockschaltbild den Teil der erfindungsgemäßen Vorrichtung, der eine individuelle skalierbare Audiosignal-Verzögerung auf der Basis der berechneten individuellen Laufzeitdifferenzen bewirkt;

Fig. 4a zeigt als Blockschaltbild eine Schaltung zur Durchführung einer dynamischen Binauralsynthese;

Fig. 4b zeigt das Blockschaltbild nach Fig. 4a mit erläuternder Beschriftung einzelner Schaltungs- bzw. Funktionsblöcke;

Fig. 5a zeigt als Blockschaltbild eine erfindungsgemäß ausgestaltete Schaltung zur Durchführung einer individualisierten dynamischen Binauralsynthese und zum Einsatz in der erfindungsgemäßen Vorrichtung; und

Fig. 5b zeigt das Blockschaltbild nach Fig. 5a mit erläuternder Beschriftung einzelner Schaltungs- bzw. Funktionsblöcke;

[0028] Die Fig. 1 zeigt das Ablaufdiagramm für ein erfindungsgemäßes Verfahren 100, das mehrere Schritte aufweist, die im wesentlichen folgendes beinhalten:

In einem ersten Schritt 110 werden aus normalen kopfbezogenen Übertragungsfunktionen, wie z.B. BRIRs, laufzeitbefreite Übertragungsfunktionen gewonnen sowie interaurale Laufzeiten extrahiert und in Differenzen umgerechnet. Dies wird noch näher anhand der Fig. 2 beschrieben, die ein schematisches Blockschaltbild einer ersten Schaltung A darstellt.

[0029] In einem zweiten Schritt 120 wird mittels der laufzeitbefreite Übertragungsfunktionen eine Binauralsynthese durchgeführt, um laufzeitbefreite Audiosignale L' bzw. R' zu erzeugen. Dies wird noch eingehender anhand der Fig. 5a/b beschrieben. Die an sich bekannte Binauralsynthese wird noch anhand der Fig. 4a/b beschrieben.

[0030] In einem dritten Schritt 130 wird eine Echtzeit-Synthese durchgeführt, um die ersten Audiosignale individuell mit einer skalierbaren Zeitverzögerung zu beaufschlagen. Dies wird noch näher anhand der Fig. 3 sowie Fig. 5a/b beschrieben.

[0031] Zunächst soll hier der erste Schritt 110 beschrieben werden: Wie die Fig. 2 veranschaulicht umfasst die Schaltung A mehrere Funktionsblöcke 111 bis 117, die jeweils einen Teilschritt der Schrittfolge 110 (s. Fig. 1) repräsentieren. Die Fig. 2 veranschaulicht somit Vorverarbeitung der Impulsantwortdatensätze, wobei eine nahezu unhörbar genaue Extraktion der ITD aus empirischen HRTF-/BRIR-Datensätzen durch Onsetdetektion erzielt werden kann.

[0032] Zunächst werden in einem Teilschritt 111 normale Übertragungsfunktionen BRIR (alternativ auch HRIR) aus einer Datenbank ausgelesen. Dann folgt eine Überabtastung um einen Betrag, der eine mehr als gehörgenaue Extraktion der Laufzeiten aus den Impulsantworten erlaubt (z.B. 10-fach bzgl. einer üblichen Audioabtastrate von 44.1 oder 48kHz) im Block 112. Danach können in einem Block 113 dann die Onsets (Beginn der Audiosignale bzw. -daten) aufgefunden werden. Anschließend wird im Block 114 die Länge der onsetbefreiten (quasi-minimalphasigen) Impulsantwort bestimmt und in einem Block 115 als Vektor angelegt. Durch anschließende, z.B. wieder 10-fache, Unterabtastung im Block 116 ergeben sich im Block 117 laufzeitfreie Übertragungsfunktionen sowie die extrahierten Laufzeiten, die als Laufzeitdifferenzen ITD berechnet und abgespeichert werden. Zusätzlich können noch beschreibende Metadatenätze DSI hinzugefügt werden. Das anhand der Fig. 2 veranschaulichte Verfahren verwendet das Onset-Verfahren zur Bestimmung der laufzeitbefreiten Impulsantworten. Hierauf wird später noch näher eingegangen.

[0033] Zunächst sei hier angemerkt, dass zwischen den an sich bekannten filtertheoretischen Aspekten von Übertragungsfunktionen, wie HRTFs oder BRIRs, und der Erfindung folgende Zusammenhänge bestehen:

Lineare, zeitinvariante und nichtrekursive Systeme, zu denen Außenohr- und binaurale Raumübertragungsfunktionen zu zählen sind, sind laut der elementaren Filtertheorie vollständig durch ihre endliche Impulsantwort beschrieben. Im digitalen Falle handelt sich damit um FIR- (finite impulse response) Filter. Diese werden in Form von mit den Filterkoeffizienten gewichteten, akkumulierenden delaylines technisch realisiert; somit stellen sie Implementierungen der diskreten Faltungssumme dar. Die zeitliche Auflösung entspricht dabei der Abtastrate des Audiosignals (z.B. 44.1 kHz). Im Falle empirischer HRTFs oder BRIRs liegen zumeist gemischtphasige Filter $x_{\text{mix}}(n)$ vor. Sie lassen sich in drei Anteile zerlegen:

$$x_{\text{mix}}(n) = x_{\text{linear}}(n) + x_{\text{allpass}}(n) + x_{\text{minimal}}(n) \quad (\text{Formel 1})$$

[0034] Die interaurale Laufzeitdifferenz ergibt sich als Differenz der Laufzeiten der HRTF bzw. BRIR des linken und rechten Ohres. Diese Laufzeiten ergeben sich wiederum als Summe des linearphasigen (reiner Laufzeitanteil) und des allpasshaltigen (frequenzabhängige Phasengangverzerrungen ohne spektrale Verzerrungen) dem sog. Exzessphasenanteil.

$$x_{\text{exzess}}(n) = x_{\text{linear}}(n) + x_{\text{allpass}}(n) \quad (\text{Formel 2})$$

[0035] Der verbleibende minimalphasige Anteil $x_{\text{minimal}}(n)$ beschreibt allein die spektralen Veränderungen (ILD) und enthält keine absoluten oder relativen Laufzeitinformationen mehr.

[0036] Die interaurale Laufzeitdifferenz ITD ist frequenzabhängig. Der für eine korrekte Lokalisation wesentliche Anteil (< 1500Hz) ist jedoch relativ konstant und lässt sich mehr oder minder artefaktfrei extrahieren. Als Verfahren eignet sich

insbesondere die Onset-Detektion, wie es anhand der Fig. 2 veranschaulicht wird. Als Alternative eignet sich auch die Bestimmung des Exzessphasenanteils durch Nutzung der Hilberttransformation, frequenzbereichsspezifisches Phasengradienten-Matching, Maximum der interauralen Kreuzkorrelation oder die frequenzbereichsspezifische Bestimmung der interauralen Gruppenlaufzeitdifferenz.

[0037] Diese genannten Methoden unterschieden sich im Wesentlichen darin, ob sie den Allpassanteil ebenfalls extrahieren können (Hilbertverfahren) oder nicht (alle anderen). Hörbare Fehler in Folge der Vernachlässigung des Allpassanteils treten jedoch nur für wenige Einfallrichtungen des Schalls auf. Zum überwiegenden Teil sind empirische HRTFs lediglich aus Kombinationen minimalphasiger und linearphasiger Anteile gebildet. Nachweislich hörbare Fehler beim Zusammensetzen "nicht-allpassenthaltender" Laufzeiten mit analytisch perfekt minimalphasigen Spektren ergeben sich nur am contralateralen (abgewandten) Ohr bei stark lateralem Schalleinfall. Inwieweit diese Verfahren auch auf in Räumen vermessene BRIRs anwenden lassen (und nicht nur auf nachhallfreie HRTFs) wurde von den Erfindern untersucht. Dabei ergab sich, dass im Falle von BRIRs nur das Onset-Verfahren genau genug (vgl. dagegen Kreuzkorrelationsverfahren), robust genug (vgl. dagegen Phasengradientenverfahren), überhaupt anwendbar (vgl. dagegen Gruppenlaufzeitdifferenzverfahren) und klangfarbentreu (vgl. dagegen Hilbertverfahren) ist. Intrinsisch bedingt konserviert das Onsetverfahren zudem mögliche Allpassanteile der BRIRs in vorteilhafter Weise; sie gehen nicht verloren, sondern bleiben in den lauffreien Spektren erhalten, die daher hier als quasi-minimalphasig bezeichnet werden. In formalen und kriterienfreien Hörversuchen wurde bestätigt, dass bei der Resynthese der extrahierten Laufzeiten und den quasi-minimalphasigen Spektren auch bei kontralateralem Schalleinfall erwartungsgemäß keine Lokalisationsfehler auftreten. Weitere formale Hörversuche zeigten, dass die aufgrund der Hilberttransformation auftretenden Veränderungen der Nachhallstruktur in jedem Falle hörbar sind.

[0038] Es sind zudem geometrisch-modellbasierte Ansätze bekannt, die aus Angaben anthropometrischer Daten des Kopfes die ITD winkelabhängig berechnen. Diese Methoden weisen subjektiv jedoch Mängel auf und haben den wesentlichen Nachteil, dass für eine modellbasiert zu generierende ITD die Schalleinfallrichtung, d.h. die virtuellen Quellpositionen, auf denen empirische HRTF- oder BRIR-Datensätze basieren, bekannt sein müssen. Bei HRTFs ist dies ein banales Problem, da eine HRTF ja genau den Schalleinfall für eine spezifische Richtung, bzw. einen spezifischen Einfallswinkel beschreibt. Bei der Vermessung von Schallfeldern in Räumen mit beweglichen Kunstköpfen sind jedoch beliebige Quellpositionen denkbar (z.B. bei simulierte Konzertsituationen). Diese sind aus den Datensätzen nicht mehr zwingend funktional basiert extrahierbar. Die an sich bekannte modellbasierte ITD-Genese ist also eher nur für parametrische, numerisch-modellhafte Ansätze der virtuellen Akustik und weniger für in-situ vermessen BRIRs geeignet.

[0039] Gemäß der Erfindung werden die kopfbezogenen Übertragungsfunktionen von den Laufzeiten befreit. Nach der Trennung von quasi-minimalphasigem Spektrum und Laufzeit können beide dann getrennt der Resynthese zugeführt werden, wobei eine skalierbare und auf das Individuum angepasste Resynthese durchgeführt werden kann (s. Fig. 3 und Fig. 5a/b). Vorteile hierbei sind Latenzreduktion und eine Verkürzung der vorzuhaltenden HRTFs (wenn diese, wie bisher implizit angenommen als FIR-Filter ausgeführt sind) um eben die extrahierte Laufzeit. Prinzipiell lassen sich die HRTFs auch als IIR-Filter (infinite impulse response) - entweder durch Modellierung/Schätzung von gemessenen HRTFs die zwangsweise immer als FIR-Filter (finite impulse response) vorliegen, oder durch parametrische Modellierung wesentlicher Features - generieren.

[0040] Das hier beschriebene Verfahren ist prinzipiell auf in beliebiger Weise erzeugte kopfbezogene Übertragungsfunktionen anwendbar. Sie wird hier aber am Beispiel von als FIR-Filtern vorliegenden empirischen HRTF-/BRIR-Datensätzen beschrieben, wie sie z.B. mit dem eingangs genannten Roboter (HATS FABIAN, siehe Lindau et al. 2007) vermessen wurden.

[0041] In Bezug auf die in der Fig. 2 veranschaulichten Extraktion aus bestehenden Übertragungsfunktionen BRIR/HRIR ist zu sagen, dass diese mittels des besagten Roboters FABIAN gewonnen werden können:

Der von den Erfindern eingesetzte Roboter FABIAN besitzt Bewegungsmöglichkeiten in 7 Freiheitsgraden (4 rotatorische, 3 translatorische). Der durch Softwaresteuerung automatisierbare Roboter besteht aus einem mit Außenohren abgeformtem Kunstkopf und generisch nachempfundenem menschlichem Körper. Der Roboter dient zum Erreichen natürlicher Schallfeldbeeinflussung (Beugung, Abschattung, Reflektion) als Resultat am Ort der eigentlichen Messung des Schallfelds, der beiden Mikrofonmembranen am geblockten Ohrkanal.

[0042] Die im Folgenden beschriebenen Algorithmen der Erfindung beziehen sich exemplarisch auf die Verwendung von BRIR-Datensätzen, die mit einem solchen Roboter gewonnen werden können. Das Verfahren ist aber nicht auf diese Datensätze beschränkt, sondern auf jegliche Auralisationsverfahren anwendbar, die räumliche Schalllokalisation durch Filterung kopfbezogener Impulsantworten mit Audiosignalen realisieren.

[0043] Das Verfahren erreicht eine wesentliche Verbesserung der auf empirischen HRTF-/BRIR-Datensätzen basierenden dynamischen Binauralsynthese und bietet insbesondere folgende Vorteile:

- i) eine nahezu unhörbar genaue Extraktion und

ii) eine ebenfalls unhörbare individuelle, echtzeitfähige Rekonstruktion und Manipulation (frequenzunabhängige Skalierung) der interauralen Laufzeitdifferenz.

[0044] Das Verfahren 100 (s. Fig. 1) umfasst dazu in einen Vorverarbeitungs-Schritt 110 (s. Fig. 1 sowie auch Teilblöcke in Fig. 2) und einen Echtzeit-Resynthese-Schritt 130 (s. Fig. 1 sowie auch Fig. 3), der einem dynamischen Binauralsynthese-Algorithmus bzw. Schritt 120 (s. Fig. 1 sowie auch Fig. 4a/b) nachgeschaltet wird (siehe Fig. 5a/b).

[0045] Zur Funktionsweise und zum Aufbau der erfindungsgemäßen Vorrichtung (s. Fig. 5a/b) sowie des von ihr ausgeführten Verfahrens (s. insbes. Fig. 1) ist zu sagen:

Zunächst erfolgt im Rahmen des Schritts 110 die Vorverarbeitung der Impulsantwortdatensätze. Das vorzuschlagende Verfahren setzt dazu bei der Extraktion der interauralen Laufzeit aus generischen HRTF-/BRIR-Datensätzen an. Die Gewinnung der BRIR-Datensätze (Block 111 in Fig. 2) erfolgt vorzugsweise mittels des zuvor beschriebenen Roboters in realen Räumen für beliebige Anzahlen von elektroakustischen Quellen und unter Verwendung eines fein aufgelöster Kopfbewegungsrasters innerhalb physiologisch sinnvoller Drehbereiche.

[0046] Liegen die empirischen Datensätze vor, so werden zunächst die Laufzeiten extrahiert (insbes. Block 113) und die interauralen Laufzeitdifferenz ITD berechnet. Dabei sei die ITD definiert als:

$$ITD = | TL - TR | \quad (\text{Formel 3})$$

wobei $TL = t_{\text{lauf}}(\text{linkes ohr})$ und $TR = t_{\text{lauf}}(\text{rechtes ohr})$

[0047] Wie bereits erwähnt, erweist sich das Onsetverfahren unter den Methoden zur Extraktion als das am besten geeignete Verfahren. Die anderen Verfahren zeigten geringere Robustheit bei empirischen Datensätzen binauraler Raumimpulsantworten. Das Hilbertverfahren erscheint zudem ungeeignet zu sein, da es aufgrund der inhärenten Energiekompaktierung in Richtung des Beginns der Impulsantwort die enthaltene Reflexionsstruktur des Raumschallfelds in evtl. hörbarem Ausmaße verändert. Dies konnten die Erfinder anhand von auditorisch-modellhaften numerischen Vorversuchen und durch einen formalen Hörversuch plausibilisieren. Das Onset-verfahren wird so eingestellt, dass das natürliche Messgrundrauschen (typ. ca. -50 bis -90 dB relativ zum Betrags-Maximalwert der Impulsantwort) bestimmt wird und dann ein Schwellwert deutlich oberhalb davon gewählt wird (z.B. um 15 dB höher, also -35dB_{rel}). Der Impulsantwortdatensatz wird dann maschinell und jeweils vom Beginn der Impulsantworten an durchsucht und anhand der Zeitpunkte des Überschreitens des Schwellwertkriteriums in linken und rechten Kanal der HRTF/BRIR durch Differenzbildung die ITD berechnet (siehe Formel 3). Die Laufzeiten werden entfernt und die nun laufzeitfreien Impulsantworten erneut gespeichert (Block 117). Das Onsetverfahren wird auf das 10-fach überabgetastete Zeitsignal angewandt, um somit eine Diskretisierung der ITD in 2.3µs-Schritten, zu erhalten. Diese Auflösung entspricht in etwa einem Fünftel des gerade wahrnehmbaren ITD-Schwellwerts. Die so extrahierte Laufzeit bzw. Laufzeitdifferenz ITD kann nun in einem textbasierten Listenformat (z.B. *.txt, *.csv, *.xml) maschinenlesbar dem laufzeitbefreiten Datensatz beigelegt werden.

[0048] Die anhand der Fig. 2 veranschaulichte Vorverarbeitungsprozedur umfasst somit folgende Teilschritte:

111: Impulsantwort einlesen (*.wav file)

112: Impulsantwort 10 x überabtasten

113: Finde Onset nach Pegel-Kriterium; Schreibe Matrix der Onsetindizes für linken und rechten Kanal; Berechne ITD aus Differenz der Onsetindizes; Korrigiere Berechnungsergebnis um Überabtastfaktor; Finde maximale Laufzeit im Datensatz

114: Bestimme neue Impulsantwortlänge, $\text{new_len} = \text{size_IR} - \text{max_time_of_flight}$ (ergibt Länge der kürzestmöglichen IR im überabgetasteten Bereich, dient als Ziellänge für alle im Folgenden extrahierten IRs)

115: Extrahiere aus Impulsantwort Vektor der Länge new_len beginnend bei Onsetindizes

116: Extrahierte Impulsantwort 10 x unterabtasten

117: Speichere IR als neue *.wav Datei

[0049] Außerdem: Speichere ITDs in maschinenlesbarer Datei; schreibe auch eine datensatzbeschreibende *.xml Datei DSI (mit Informationen über Start und Endpunkte, Auflösungen, für Azimut und Elevation).

[0050] Diese Verarbeitung wird in einer ersten Schaltung A durchgeführt, die Teil der Vorrichtung ist (s. Fig. 5a/b).

[0051] Eine weitere Schaltung B mit Funktionsblöcken BA und BB (s. Fig. 4a/b) wird in einer abgewandelten Form (mit Funktionsblock BB*) ebenfalls in der Vorrichtung (s. Fig. 5a/b) eingesetzt und dient einer individuellen, adaptiven Echtzeit-Resynthese, die nachfolgend noch näher beschrieben wird:

Ausgangspunkt bzw. Vorbedingung für die individualisierte binaurale Audiowiedergabe ist das Vorhandensein eines Echtzeitfaltungsalgorithmus, wie er an sich bekannt ist und anhand der Fig. 4a/b veranschaulicht wird. Dieser sollte die Faltung blockbasiert und individuell für jede einzelne zu auralisierende Schallquelle vornehmen können. Dabei unterscheidet sich der technische Vorgang der schnellen Echtzeitfaltung von laufzeitbehafteten und laufzeitlosen BRIRs nicht, allerdings ergeben sich aufgrund der Datenaufbereitung spezielle klangliche Vorteile, die weiter unten noch beschrieben werden.

[0052] Zur Abwandlung des Faltungsprozess bzw. zur Laufzeitindividualisierung wird vorzugsweise auf eine sogenannte "jack audio server"-Architektur zurückgegriffen und damit ein eigenständiges PlugIn realisiert. Funktionsgleiche Lösungen können z.B. auch aus Anordnungen spezieller DSP-Hardware oder mittels Verfahren im Rahmen der VST-PlugIn-Architektur (d.h. auf der VST-Schnittstelle basierend) realisiert werden. Nach dem Start der rechnergesteuerten Vorrichtung (z.B. PC) liest diese eine Konfigurationsdatei, eine Datensatzbeschreibungsdatei und danach die textbasierte Liste der ITDs des aktuell vom klassischen Faltungsprozess auralisierten BRIR-Datensatzes ein. Dann meldet es sich beim Audioserver an und stellt je Instanz (d.h. je individuell anpassbar zu auralisierender Audioquelle) zwei Ein- und Ausgänge zu Verfügung (für linkes und rechtes binaurales Signal des Faltungsprozesses). Die Eingänge können dann in PlugIn-Manier mit den Ausgängen des Faltungsprozesses verbunden werden (siehe auch Fig. 5a/b). Für mehrere zu auralisierende Quellen sind jeweils ein eigener Faltungsprozess und ein Prozess zur Laufzeitindividualisierung zu starten. Der Prozess zur Laufzeitindividualisierung kopiert nun blockweise das Ausgangsergebnis des Faltungsprozesses (binaurale Zeitsignale ohne Laufzeitdifferenzen, siehe L' und R' in Fig. 5a/b) in seine eigenen Eingangspuffer. Damit die Grundlatenz in dieser seriellen Kaskadierung nicht über die Gebühr ansteigt haben sich Blockgrößen von 128-512 samples für beide Prozesse bewährt.

[0053] Die fehlende Laufzeitdifferenz kann nun als kopfpositionsspezifische Verzögerungszeit VDL und ohne hörbare Artefakte in einen der beiden Audiokanäle eingefügt werden. Die geschieht hier z.B. für L' im Block VDL/SCR bzw. S, wobei der Block VDL/SCR die dynamische Anpassung der Laufzeit bei Kopfbewegung darstellt und der Block S die statische Anpassung, die sich bei fester Kopfposition dann ergibt bzw. einstellt. Für die Anpassung liest der Individualisierungsprozess die Kopfbewegungsdaten als Datenstrom des Headtrackers HTDAT (z.B. via ein IP-basiertes Transportprotokoll) mit, die dieser wie bisher auch an den Faltungsprozess sendet. Bei der Initialisierung des Individualisierungsprozesses muss dieser die, durch die initiale Kopfposition bestimmte, erste wirksame interaurale Laufzeitdifferenz durch um einen fixen Betrag zeitverzögertes oder vorgegreifendes Abspielen realisieren. Der Einfachheit wg. kann die erste Zeitdifferenz zu 0, bzw. der erste Konversionsfaktor zu 1, angenommen werden. Anschließend werden kopfpositionsbedingte Veränderungen der interauralen Laufzeitdifferenz wie folgt erreicht: Je verarbeitetem Audioblock wird die aktuelle Kopfposition und damit eine Veränderung der Laufzeitdifferenzen geprüft. Wurden Veränderungen der Laufzeitdifferenzen detektiert, wird deren Ausmaß durch Differenzbildung ($\Delta \text{ITD} = \text{neue ITD} - \text{alte ITD}$) bestimmt. Aus der fraktionalen Verhältnisbildung der Audioblocklänge mit der um das Ausmaß der Veränderung der Laufzeitdifferenzen (ΔITD) korrigierten Audioblocklänge ergibt sich ein Verhältnis, welches als Konversionsfaktor in einem Algorithmus zur Echtzeitabtastratenkonversion in höchster Audioqualität verwendet werden kann. Dieser erreicht die Anpassung der veränderten Verzögerungswerte durch ein um den Konversionsfaktor beschleunigtes oder verzögertes Abspielen, in dem es aus dem vorliegenden Signal neue Abtastwerte an anderen, einer höheren oder niedrigeren Abtastrate entsprechenden, Zeitpunkten interpoliert und diese stattdessen ausgibt.

[0054] Mit Hilfe einer Abtastratenkonversion SRC können korrekte Amplitudenwerte zu beliebigen kontinuierlichen Zeitpunkten aus den Abtastzeitpunkten digitaler Zeitsignale gefunden werden. An sich bekannt ist auch die sog. fraktionale Abtastratenkonversion (siehe Smith III, J.O. (2002): Digital Audio Resampling Home Page. Web published at <http://www-ccrma.stanford.edu/~jos/resample/>, Stanford: Center for Computer Research in Music and Acoustics (CCRMA), Stanford University). Nach Shannons Theorem der perfekten Rekonstruktion des Zeitverlaufs abgetasteter Signale durch ebenso perfekte Tiefpassfilter (Antialiasingfilter) nutzt das Verfahren der bandbeschränkten Interpolation hierzu eine Tabelle interpolierter quasikontinuierlicher Filterkoeffizienten, um die jeweilige Tiefpassimpulsantwort effektiv in kontinuierlicher Zeitsignalfom verfügbar zu machen. Eine einzelne vorberechnete Filtertabelle genügt bereits für alle Interpolationszeiten und SRC-Verhältnisse. Je nach Abtastratenkonversionsverhältnis $f_{s_{\text{neu}}}/f_{s_{\text{alt}}}$ (d.h. der Konversionsfaktor ist <1 oder >1) ist ein Filter mit niedrigerer oder höherer Eckfrequenz aus der Tabelle auszulesen. Das wiederum ist möglich, da die Filterimpulsantwort quasikontinuierlich, d.h. mit sehr hoher Abtastgenauigkeit in sehr großer Länge tabelliert ist. Für jeden aktuell zu resampelnden Block von Ausgangssamples - bei Blöcken von sehr kurzer Dauer ist variable Konversion somit annähernd in "Echtzeit" möglich - wird das Interpolationsergebnis an den neuen Abtastzeitpunkten durch Bildung der Faltungssumme der Zeitsignalsamples mit der neuen Tiefpassimpulsantwort berechnet. Je nach benötigter Güte (d.h. Bandbreite und Sperrdämpfung des Interpolators) kann die Tabellengröße des Interpolatortiefpasses durch geeignete Formeln abgeschätzt werden. Benutzt wird hier ein hochqualitativer, bandbegrenzter Fliesskomma-Interpolator auf Basis einer analytisch beschriebenen sinc-Funktion mit einem worst case Signalrauschverhältnis von 97 dB bei einer Bandbreite von 97%.

[0055] Wie bereits erwähnt, kann der je zu verarbeitendem Audioblock aktuelle Konversionsfaktor jeweils mithilfe der

Differenz der zur vorherigen und zur aktuellen Kopfposition gehörenden ITD bestimmt werden. Eine jederzeit verfügbare Softwarebibliothek erlaubt eine unhörbare und kontinuierliche Änderung des Konversionsfaktors, so dass die der jeweiligen Kopfhaltung entsprechenden Laufzeitdifferenz korrekt resynthetisiert werden kann. Für die praktische Realisierung der interauralen Laufzeitdifferenz bedeutet das, dass mittels der o.g. Bibliothek zur bandbeschränkten Interpolation aus einer variablen Anzahl Eingangssamples eine immergleiche - der zu Grunde liegenden Audioblocklänge entsprechende - Anzahl neuer Ausgangssamples erzeugt wird. Je nach Abtastratenkonversionsverhältnis $f_{s_{\text{neu}}}/f_{s_{\text{alt}}}$ (<1 oder >1) benötigt die Abtastratenkonversion dafür weniger oder mehr Samples als die zu Grunde liegende aktuelle Blockgröße. Aufgrund dieser Differenz erhält man als Ergebnis - außer dem Block neu interpolierter Ausgangssamples - eine feste zeitliche Verschiebung des Einlesebereichs der Abtastratenkonversion um die Anzahl im vorhergegangenen Block von einer Abtastratenkonversion weniger oder zusätzlich benötigten Samples. In der Fig. 5a/b wird diese Verschiebung durch den Block S als implizite, feste und samplegenaue Verzögerung dargestellt, deren aktueller Wert sich aus allen vorhergegangenen Laufzeitdifferenzveränderungen ergibt.

[0056] Wurde während zweier Audioblöcke keine Veränderungen der interauralen Laufzeitdifferenz detektiert, bleibt dieser feste Verzögerungswert (statisches Delay S) einfach erhalten und realisiert so weiterhin die korrekte interaurale Laufzeitdifferenz.

[0057] Im Vorgriff auf die im Folgenden beschriebenen individuelle Skalierung des Konversionsfaktors, ist zudem zu betonen, dass diese feste Verzögerung ebenfalls korrekt individuell skaliert ist, da sie nur aus der Akkumulation von Rückgabewerten der bereits individuell skalierten Abtastratenkonversion resultiert.

[0058] Ein bedeutender Vorteil der Erfindung besteht nun darin, dass die Laufzeitdifferenz durch einfache Skalierung des jeweils aktuellen Konversionsfaktors individuell angepasst werden kann. Dazu wird die ITD, z.B. vor der Berechnung des Konversionsfaktors, einfach mit einem festen Wert skaliert (praktische Werte liegen zw. 50-150%). Damit ergibt sich zwischen den Ausgangssignalen L^* und R^* eine individuell angepasste ITD*. Bereits mit einer solchen frequenzunabhängigen Skalierung der ITD konnte eine perzeptiv deutlich verbesserte Anpassung an das jeweilige Individuum erzielt werden. Zudem ist diese Skalierung der ITD, welche näherungsweise einer nachträglichen Skalierung des Kopfdurchmessers entspricht, in Echtzeit regelbar. Wie ein optimaler, individueller Skalierungsfaktor gefunden werden kann, soll anhand der folgenden zwei Beispiele verdeutlicht werden:

Der Ansatz für das erste Beispiel beruht auf der Einrichtung des Faktors durch die Versuchsperson selbst: Da der Korrekturwert in Echtzeit einregelbar ist, kann dieser zu Beginn der virtuellen akustischen Anwendung mit Hilfe einer geeigneten akustischen Szene (z.B. Darstellung einer einzelnen frontalen Schallquelle, die kurze gut lokalisierbare Rauschpulse abgibt) eingestellt werden. Dazu verstellt die Person den Korrekturwert beginnend von einem zunächst sehr hoch gewählten Startwert (z.B. 150% der ITD) in negativer Richtung, bis sich eine korrekte Lokalisationsempfindung, gekennzeichnet durch eine stabile frontale Quellenlokalisierung, d.h. vor allem ohne die den Kopfbewegungen folgende Interaktionsbewegungen einstellt. Der zu Beginn hoch gewählte Startwert macht das zu minimierende Artefakt besonders gut hörbar.

[0059] Ein weiterer Ansatz, der als zweites Beispiel dienen soll, beruht auf einer Vorhersage des individuellen Korrekturwerts auf Basis eines anthropometrischen Maßes. Es wurden Voruntersuchungen mit mehreren Versuchspersonen durchgeführt. Dabei wurde ein Hörversuch entsprechend der oben genannten akustischen Szenerie durchgeführt; zusätzlich zur virtuellen Quelle konnten die Versuchspersonen jedoch auch die reale Schallquelle hören. Das Ziel war es, den Konversionsfaktor der Simulation so einzustellen, dass sich beim Umschalten zw. Simulation und Realität eine möglichst gute Übereinstimmung ergab. Dieser Versuch wurde je Versuchsperson 10-mal wiederholt. Zusätzlich wurden vier als geeignete erachtete Maße des Kopfes von jeder Person erhoben. Anschließend wurde die Vorhersage der Mittelwerte des individuellen Skalierungsfaktors aus den Kopfabmessungen durch multiple lineare Regression überprüft. Diese ergab, dass allein die Intertragusdistanz als Maß des Kopfdurchmessers, gemessen mit einem Tastzirkel in Millimetern von der incisura anterior des linken zu der des rechten Ohres genügt, um den Konversionsfaktor vorherzusagen. Grafische Darstellungen der Untersuchungen zeigten die Regressionsgerade und das 95%-ige Konfidenzintervall des linearen Modells sowie die individuellen Mittelwertdaten aus der Eichstichprobe. Das 95%-KI zeigte dabei an, dass der Konversionsfaktor mit einer Genauigkeit von $\pm 1.25\%$ (mittlerer Modellbereich) bis $\pm 2.5\%$ (äußere Modellbereiche) vorhergesagt werden kann.

[0060] Die Variabilität des Skalierungsfaktors innerhalb der Stichprobe betrug dabei $\pm 4\%$, d.h. dass im nichtindividualisierten Fall ein worst-case Fehler von bis zu 8% der ITD hätte auftreten können. Die individuelle Vorhersage halbiert bzw. viertelt (im günstigsten Fall) diesen Fehler. Ein Fehler von 1.25% der ITD entspricht dabei einem Lokalisationsfehler von etwas über 1° und liegt damit (wieder: im günstigsten Fall) schon nahezu im Bereich der gerade wahrnehmbaren Lokalisationsänderung. Die Regressionsformel zur Vorhersage des individuellen Korrekturfaktors lautet dabei:

$$\text{Scale} = 0.00304 * \text{Intertragusdistanz [mm]} + 0.5792$$

[0061] Diese Regressionsgleichung liefert zwar eine Vorhersage des individuellen Korrekturfaktors, sie ist in dieser Form jedoch nur auf Datensätze anwendbar, die mit dem FABIAN HATS Roboter vermessen wurde. Eine Möglichkeit zur Verallgemeinerung auf beliebige HRTF/BRIR-Datensätze ergäbe sich durch eine Korrektur des Ergebnisses der o. g. Formel durch Multiplikation mit dem Kopfdurchmesser Verhältnis Intertragusdistanz Fremd/Intertragusdistanz Fabian. Diese Korrektur setzt natürlich voraus, dass die Intertragusdistanz des fremden Datensatzes bekannt ist.

[0062] Das Prinzip der erfindungsgemäßen Resyntheseprozedur wird anhand der Fig. 3 veranschaulicht:

a) Initialisierung

- i) Lese *.xml Konfigurationsdatei
- ii) Lese Datensatzdefinition aus *.xml Beschreibungsdatei
- iii) Lege 1 oder 2-dimensionales Array durch Lesen der ITDs aus maschinenlesbarer Datei an
- iv) Setze initialen Skalierungsfaktor der ITD = 1
- v) Initialisiere Ringpuffer des SRC-Prozesses
- vi) Starte OSC-Server für Empfang der Kopfbewegungsdaten
- vii) Initialisiere jack client.
- viii) Starte Echtzeitbedienungsmodi (OSC, GUI, Keyboard)

b) Echtzeitprozess

i) Kopiere Eingangssamples für Links und Rechts in Ringpuffer

- a. Lese ITD bzgl. Kopfposition laut head tracking
- b. Skaliere ITD in Echtzeit, wenn gewollt ($\text{ITD}_{\text{new}} = \text{ITD} * \text{scaling_factor}$)
- c. Rechne ITD fraktional in Samples um ($\text{delay_new} = \text{ITD}_{\text{new}} / \text{sample_size}$)
- d. Bestimme ITD-Veränderung seit letzter Berechnung ($\text{delta_delay} = \text{delay_new} - \text{delay_old}$)
- e. Bestimme fraktionalen SRC-Faktor ($\text{ratio} = \text{output_samples} / (\text{input_samples} - \text{delta_delay})$)

- ii) Übergebe an SRC-Prozess: Faktor, Eingangs- und Ausgangspuffer für linken Kanal.
- iii) Speichere ungenutzte Samples im Ringpuffer
- iv) Berechne $\text{input_buffer_length}$ des nächsten Zyklus' ($\text{input_buffer_length} = \text{output_samples} + \text{ungenutzte Samples}$)
- v) Gebe Ergebnis aus: linker Kanal aus SRC-Prozess; rechter Kanal aus Ringpuffer

[0063] Durch die Erfindung ergeben sich u.a. folgende besondere Vorteile und Verbesserungen gegenüber dem Stand der Technik:

- 1) Durch das Erreichen einer deutlich genaueren Lokalisation für alle Kopfpositionen unter Verwendung generischer Datensätze sorgt die individuell skalierte ITD wesentlich für die Erhöhung der Plausibilität einer binauralen Simulation.
- 2) Die sich aufgrund der Extraktion der ITD ergebende quasi-minimalphasige Überblendung hat deutlich hörbare auditive Vorteile, da bei der Überblendung quasiminimalphasiger Spektren so gut wie keine Überblendungsartefakte mehr auftreten.
- 3) Ein weiterer wesentlicher Vorteil ergibt sich aus der Quasi-Minimalphasigkeit der Spektren. Sie bildet eine Grundlage für alle effizienten spektralen & temporalen Kompressions- und Interpolationsverfahren für HRTF-/BRIR-Datensätze und kann z.B. auch auf beschränkten Plattformen, wie z.B. Mobilfunkgeräten, eingesetzt werden.
- 4) Die Operationalisierung der ITD erlaubt eine getrennte Interpolation von ILD- und ITD-Features. So kann z.B. die perzeptiv wichtigere ITD durchaus mit einer sehr feinen Auflösung resynthetisiert werden, während die minimalphasigen Spektren mit relativ grober Auflösung (z.B. 5 Grad) verarbeitet werden, was wiederum auch den Einsatz in Mobilfunkgeräten erleichtert.
- 5) Die Entfernung der akustischen Laufzeit aus allen BRIR-Datensätzen führt zu Latenzminimierung und Speicher-

platzreduktion, so dass auch hier sich insbesondere der Einsatz bei Mobilfunkgeräten anbietet.

6) Die Implementierung der ITD durch Resampling führt zu einer "physikalischeren" Modellierung des Schallfelds, da sich aufgrund des verlangsamten bzw. beschleunigten Abspielens innerhalb des SRC-Prozesses eine (zumindest für den Direktschall korrekte) Imitation des Dopplereffekts einstellt.

7) Durch die Aufspaltung der Faltung in Einzelprozesse je Quelle wird die Voraussetzung für eine parallelisierte Berechnung (Erhöhung der Berechnungseffizienz) geschaffen.

[0064] Die Erfindung ermöglicht das Erreichen zahlreicher Verbesserungen, wie z.B. den Vorteil der fehlerfreien Überblendung, die Latenzminimierung, den Dopplereffekt. Hervorzuheben sind auch die Wahl der bandbegrenzten Interpolation zur Abstratenkonversion bei der Generierung der variablen ITD, die Echtzeitfähigkeit der ITD-Manipulation, die unhörbare Extraktionsmethode und die PlugIn-Architektur. Ebenso sind wichtige Aspekte, wie z.B. die effektive und getrennte Reduktion und Interpolation von ITD und ILD (d.h. HRTF/BRIR-Spektren), hervorzuheben.

[0065] Die individuelle Anpassung der laufzeitbefreiten ITDs an das jeweilige Individuum kann durch Skalierung des jeweils aktuellen Konversionsfaktors im Rahmen einer Abstratenkonversion erfolgen, die bei Änderung der Kopfposition eine dynamische Anpassung ermöglicht. Sofern ein statischer Zustand sich einstellt, d.h. die Kopfposition sich nicht ändert, stellt sich auch eine statische Anpassung der ITDs (automatisch) ein. Oder die individuelle Anpassung erfolgt durch Skalierung der ITDs, die zu der vorherigen und aktuellen Kopfposition gehören, bevor der (unskalierte) Konversionsfaktor berechnet wird.

[0066] Mit anderen Worten: Die Individualisierung und Anpassung der sich zwischen den Ausgangssignalen L^* und R^* ergebenden Laufzeit-Differenz ITD^* kann erreicht werden durch Skalierung des jeweils aktuellen Konversionsfaktors (Anwendung des Faktors ISF auf den Block VDL/SRC; siehe Fig. 5a/b). Alternativ kann die Individualisierung und Anpassung der sich zwischen den Ausgangssignalen L^* und R^* ergebenden LaufzeitDifferenz ITD^* erreicht werden durch Skalierung der zu den laufzeitbefreiten kopfbezogenen Übertragungsfunktionen ($xBIR^*$) korrespondierenden Laufzeit-Differenzen (ITD).

[0067] Die Erfindung ist in vielen Anwendungen einsetzbar. Durch das vorgeschlagene Verfahren kann eine wesentliche Verbesserung aller bisher existierender Echtzeit-Anwendungen zur binauralen (Raum-)Akustiksimulation (3D Virtual Auditory Displays im Allgemeinen, räumliche Akustik in Computerspielen, Virtuelle Chaträume, Binaurale Guidance & Alerting Systeme, binaurale Walkthroughs z. B. durch virtuelle Architektur oder durch multimodale Media-shows) erzielt werden. Mögliche Anwendungen sind die nachträgliche Ergänzung kommerzieller 3D Audio APIs. Auch die Qualität zukünftig denkbarer rich media Anwendungen die empfängerseitig einen Multimedia-PC voraussetzen wie z. B. binaurales Teleconferencing via VoIP oder binaurales Streaming von Livekonzerten kann wesentlich an wahrnehmbarer Qualität gewinnen.

Patentansprüche

1. Verfahren (100) zur Erzeugung individuell anpassbarer binauraler Audiosignale (L^* , R^*), mit folgenden Schritten:

aus kopfbezogenen Übertragungsfunktionen ($xRIR$), die verschiedenen Kopfpositionen (1...k) zugeordnet sind, werden mittels Extraktion von interauralen Laufzeiten (TL, TR) laufzeitbefreite kopfbezogene Übertragungsfunktionen ($xRIR^*$) und dazu korrespondierende Laufzeit-Differenzen (ITD) berechnet (Schritt 110);

mittels einer Binauralsynthese werden laufzeitbefreite Audiosignale (L' , R') erzeugt, die sich jeweils auf einen binauralen Signalfad beziehen, wobei in Abhängigkeit von Kopfpositions-Daten (HTDAT) die laufzeitbefreiten kopfbezogenen Übertragungsfunktionen ($xRIR^*$) für eine aktuelle Kopfposition (1...k) verwendet werden (Schritt 120); **dadurch gekennzeichnet,**

dass aus den laufzeitbefreiten Audiosignalen (L' , R') die individuell anpassbaren binauralen Audiosignale (L^* , R^*) erzeugt werden, indem zumindest für einen Signalfad das erzeugte laufzeitbefreite Audiosignal (L') mit einer zeitlichen Verzögerung (VDL) in Abhängigkeit von der für die aktuelle Kopfposition (1...k) berechneten Laufzeit-Differenz (ITD) und in Abhängigkeit eines individuellen Skalierungsfaktors (ISF) beaufschlagt wird (Schritt 130),

dass bei einer auftretenden Änderung von einer vorherigen Kopfposition ($k=1$) in die aktuelle Kopfposition ($k=2$) die zeitliche Verzögerung (VDL) zwischen einem ersten Wert und einem zweiten Wert adaptiv mittels einer Abstratenkonversion (SRC) verändert wird, wobei die Abstratenkonversion einen Konversionsfaktor verwendet um das laufzeitbefreite Audiosignal um den Konversionsfaktor beschleunigt oder verzögert abzuspielen (Schritt 130), und

dass der für die Abstratenkonversion (SRC) verwendete Konversionsfaktor entsprechend der Änderung von der zu der vorherigen und zu der aktuellen Kopfposition gehörenden zeitlichen Verzögerung ermittelt wird (Schritt 130).

- 5 **2.** Verfahren (100) nach Anspruch 1, **dadurch gekennzeichnet, dass** die zeitliche Verzögerung (VDL) einem Wert entspricht, der aus der Wichtung der für die aktuelle Kopfposition (1...k) berechneten Laufzeit-Differenz (ITD) und dem individuellen Skalierungsfaktor (ISF) berechnet wird (Schritt 130).
- 10 **3.** Verfahren (100) nach Anspruch 1, **dadurch gekennzeichnet, dass** eine zwischen individuell anpassbaren binauralen Audiosignalen (L*, R*) sich ergebende Laufzeit-Differenz (ITD*) durch Skalierung des aktuellen Konversionsfaktors individuell angepasst wird (Schritt 130).
- 15 **4.** Verfahren (100) nach Anspruch 3, **dadurch gekennzeichnet, dass** sich die ergebende Laufzeit-Differenz (ITD*) durch Skalierung der zu den laufzeitbefreiten kopfbezogenen Übertragungsfunktionen (xBIR*) korrespondierenden Laufzeit-Differenzen (ITD) individuell angepasst wird.
- 20 **5.** Verfahren (100) nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** die laufzeitbefreiten kopfbezogenen Übertragungsfunktionen (xRIR*) für eine Vielzahl von Kopfpositionen (1...k) und für eine Vielzahl von Audiosignal-Quellen (1...n) berechnet werden (Schritt 110).
- 25 **6.** Verfahren (100) nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** die interauralen Laufzeiten (TL, TR) aus den kopfbezogenen Übertragungsfunktionen (xRIR) mittels eines der folgenden Verfahren extrahiert werden: Onset-Verfahren, interaurales Kreuzkorrelationsverfahren, frequenzspezifische Gruppenlaufzeit-Differenzbildung, frequenzbereichsspezifisches Phasengradienten-Matching oder Exzessphasenbestimmung durch Hilbert-Transformation (Schritt 110).
- 30 **7.** Verfahren (100) nach Anspruch 6, **dadurch gekennzeichnet, dass** das Onset-Verfahren so eingestellt wird, dass ein natürliches Messgrundrauschen bestimmt wird und dann ein Schwellwert oberhalb davon, der niedriger als das Betrags-Maximum der durch die jeweilige kopfbezogene Übertragungsfunktion bestimmten Impulsantwort ist, bestimmt wird (Schritt 110).
- 35 **8.** Verfahren (100) nach Anspruch 7, **dadurch gekennzeichnet, dass** der Schwellwert um mindestens 10 dB oberhalb des natürlichen Messgrundrauschen bestimmt und/oder das der Schwellwert um mindestens 10 dB niedriger als das Betrags-Maximum der durch die jeweilige kopfbezogene Übertragungsfunktion bestimmten Impulsantwort ist (Schritt 110).
- 40 **9.** Verfahren (100) nach Anspruch 8, **dadurch gekennzeichnet, dass** der Schwellwert im mittleren Bereich einer Messdynamik liegt (Schritt 110).
- 45 **10.** Verfahren (100) nach Anspruch 6, **dadurch gekennzeichnet, dass** das Onset-Verfahren so eingestellt wird, dass ein prozentualer Wert vom Betrags-Maximum der durch die jeweilige kopfbezogene Übertragungsfunktion bestimmten Impulsantwort als Onset detektiert wird, insbesondere als ein zwischen 10% und 90% liegender Wert detektiert wird (Schritt 110).
- 50 **11.** Vorrichtung zur Durchführung des Verfahrens nach einem der vorhergehenden Ansprüche, wobei die Vorrichtung zur Erzeugung räumlich wahrnehmbarer binauraler Audiosignale (L*, R*) umfasst:
 eine erste Schaltung (A), die aus kopfbezogenen Übertragungsfunktionen (xRIR), die verschiedenen Kopfpositionen (1...k) zugeordnet sind, interaurale Laufzeiten (TL, TR) extrahiert und daraus Laufzeit-Differenzen (ITD) berechnet;
 eine zweite Schaltung (BB), die mittels einer Binauralsynthese laufzeitbefreite Audiosignale (L', R') erzeugt, die sich jeweils auf einen binauralen Signalpfad beziehen, wobei diese Schaltung in Abhängigkeit von Kopfpositions-Daten (HTDAT) die laufzeitbefreiten kopfbezogenen Übertragungsfunktionen (xRIR*) für eine aktuelle Kopfposition (1...k) verwendet; **gekennzeichnet durch**
 mindestens eine dritte Schaltung (BA*; C*), die aus den laufzeitbefreiten Audiosignalen (L', R') die individuell anpassbaren binauralen Audiosignale (L*, R*) erzeugt, indem diese Schaltung zumindest für einen Signalpfad das erzeugte laufzeitbefreite Audiosignal (L') mit einer zeitlichen Verzögerung (VDL) in Abhängigkeit von der für die aktuelle Kopfposition (1...k) berechneten Laufzeit-Differenz (ITD) und in Abhängigkeit eines individuellen

Skalierungsfaktors (ISF) beaufschlagt,

wobei die mindestens eine dritte Schaltung (BA*; C*) bei einer auftretenden Änderung von einer vorherigen Kopfposition (k=1) in die aktuelle Kopfposition (k=2) die zeitliche Verzögerung (VDL) zwischen einem ersten Wert und einem zweiten Wert adaptiv mittels einer Abtastratenkonversion (SRC) verändert, wobei die Abtastratenkonversion einen Konversionsfaktor verwendet um das laufzeitbefreite Audiosignal um den Konversionsfaktor beschleunigt oder verzögert abzuspielen, und den für die Abtastratenkonversion (SRC) verwendeten Konversionsfaktor entsprechend der Änderung von der zu der vorherigen und zu der aktuellen Kopfposition gehörenden zeitlichen Verzögerung ermittelt.

Claims

1. A method (100) for generating individual adjustable binaural audio signals (L*, R*), comprising the following steps:

from head-related transfer functions (xRIR) which are assigned to various head positions (1...k) head-related transfer functions (xRIR*) being freed of transit time are calculated by means of extraction of interaural transit times (TL, TR), and corresponding transit time runtime differences (ITD) are calculated (step 110); by means of a binaural synthesis audio signals (L', R') being freed of transit time are generated, each of which relates to a binaural signal path, wherein in dependence from head position data (HTDAT) the head-related transfer functions (xRIR*) being freed of transit time are used for an actual head position (1...k) (step 120);

characterized in

that from the audio signals (L', R') being freed of transit time the individual adjustable binaural audio signals (L*, R*) are generated by loading for at least one signal path the generated audio signal (L') being freed of transit time with a time delay (VDL) in dependence from the transit time difference (ITD) being calculated for the actual head position (1...k) and in dependence from an individual scaling factor (ISF) (step 130);

that when a change of the former head position (k=1) to the actual head position (k=2) occurs, the time delay (VDL) between a first value and a second value is adaptively changed by a sampling rate conversion (SRC), wherein the sampling rate conversion applies a conversion factor to reproduce with advance or delay the audio signal being freed of transit time (step 130); and

that the conversion factor being used for the sampling rate conversion (SRC) is determined according to the time delay change corresponding to the change of the former head position (k=1) to the actual head position (k=2) (step 130).

2. The method (100) of claim 1, **characterized in that** the time delay (VDL) corresponds to a value which is calculated from a weighting of the transit time difference (ITD) being calculated for the actual head position (1...k) and of the individual scaling factor (ISF) (step 130).

3. The method (100) of claim 1, **characterized in that** a transit time difference (ITD*) arising between individual adjustable binaural audio signals (L*, R*) is adjusted individually by scaling of the actual conversion factor (step 130).

4. The method (100) of claim 3, **characterized in that** the arising transit time difference (ITD*) is individually adjusted by scaling of the transit time differences (ITD) corresponding to the head-related transfer functions (xRIR*) being freed of transit time.

5. The method (100) of one of the preceding claims, **characterized in that** the head-related transfer functions (xRIR*) being freed of transit time are calculated for a plurality of head positions (1...k) and for a plurality of audio signal sources (1...n) (step 110).

6. The method (100) of one of the preceding claims, **characterized in that** the interaural transit times (TL, TR) are extracted from the head-related transfer functions (xRIR) by means of one of the following methods: onset method, interaural cross correlation method, frequency specific group delay (transit time) difference building, frequency range specific phase gradient matching or excess phase determination by Hilbert transformation (step 110).

7. The method (100) of claim 6, **characterized in that** the onset method is applied such that a natural measuring noise floor is determined and then a threshold located above is determined which is lower than the absolute maximum of the impulse response resulting from the corresponding head-related transfer function (step 110).

8. The method (100) of claim 7, **characterized in that** the threshold is determined to be least 10 dB above the natural

measuring noise floor and/or that the threshold is determined to be least 10 dB lower than the absolute maximum of the impulse response resulting from the corresponding head-related transfer function (step 110).

9. The method (100) of claim 8, **characterized in that** the threshold is in an average range of the measuring dynamics (step 110).

10. The method (100) of claim 6, **characterized in that** the onset method is applied such that a value in percent of the absolute maximum of the impulse response resulting from the corresponding head-related transfer function is detected to be the onset, in particular as a value between 10% and 90% (step 110).

11. Device for performing the method of one of the preceding claims, wherein the device for generating binaural audio signals (L^* , R^*) for spatial perception, comprising:

a first circuit (A) for extracting from head-related transfer functions (xRIR) which are assigned to various head positions (1...k) head-related transfer functions (xRIR*) being freed of transit time, and for calculating transit time differences (ITD) thereof;

a second circuit (BB) for generating by means of a binaural synthesis audio signals (L' , R') being freed of transit time, each of which relating to a binaural signal path, wherein this circuit uses in dependence from head position data (HTDAT) the head-related transfer functions (xRIR*) being freed of transit time for an actual head position (1...k); **characterized by**

at least a third circuit (BA^* ; C^*) for generating from the audio signals (L' , R') being freed of transit time the individual adjustable binaural audio signals (L^* , R^*) by loading for at least one signal path the generated audio signal (L') being freed of transit time with a time delay (VDL) in dependence from the transit time difference (ITD) being calculated for the actual head position (1...k) and in dependence from an individual scaling factor (ISF);

wherein the at least third circuit (BA^* ; C^*), when a change of the former head position ($k=1$) to the actual head position ($k=2$) occurs, adaptively changes the time delay (VDL) between a first value and a second value by means of a sampling rate conversion (SRC), wherein the sampling rate conversion applies a conversion factor to reproduce with advance or delay the audio signal being freed of transit time; and

determines the conversion factor being used for the sampling rate conversion (SRC) according to the time delay change corresponding to the change of the former head position to the actual head position.

Revendications

1. Une méthode (100) pour générer des signaux audio binauraux individuellement adaptables (L^* , R^*), comprenant les étapes suivantes :

le calcul, à partir de fonctions de transfert de tête (xRIR) associées à diverses positions de la tête (1...k), de fonctions de transfert de tête (xRIR*) dépourvus de temps de transit grâce à l'extraction de temps de transit interaural (TL, TR) et le calcul des différences de temps de fonctionnement (ITD) de transit correspondants (étape 110);

la génération, au moyen d'une synthèse binaurale, de signaux audio (L' , R') dépourvus de temps de transit, chacun relatif à un chemin de signal binaural, dans laquelle, en fonction des données de position de tête (HTDAT), les fonctions de transfert de tête (xRIR*) dépourvues de temps de transit sont utilisées pour une position actuelle de la tête (1... k) (étape 120);

caractérisée en ce que

les signaux audio binauraux individuellement adaptables (L^* , R^*) sont générés à partir des signaux audio (L' , R') dépourvus de temps de transit en chargeant pour au moins un chemin de signal le signal audio généré (L') dépourvu de temps de transit un délai (VDL) fonction de la différence de temps de transit (ITD) calculée pour la position de tête réelle (1...k) et en fonction d'un facteur d'échelle individuel (ISF) (étape 130);

lorsque intervient un changement depuis une position précédente de tête ($k=1$) vers la position réelle de tête ($k=2$), le délai (VDL) entre une première valeur et une seconde valeur est changée de manière adaptative au moyen d'une conversion de taux d'échantillonnage (SRC), dans laquelle la conversion de taux d'échantillonnage applique un facteur de conversion pour reproduire le signal audio dépourvu de temps de transit en avance ou en retard (étape 130) ; et

le facteur de conversion utilisé pour la conversion du taux d'échantillonnage (SRC) est déterminé en fonction du changement de délai correspondant au changement de la position de tête précédente ($k=1$) vers la position

de tête actuelle ($k=2$) (étape 130).

2. La méthode (100) de la revendication 1, **caractérisée en ce que** le délai (VDL) correspond à une valeur qui est calculée à partir d'une pondération de la différence de temps de transit (ITD) calculé pour la position de tête actuelle ($1...k$) et pour le facteur d'échelle individuel (ISF) (étape 130).
3. La méthode (100) de la revendication 1, **caractérisée en ce que** une différence de temps de transit (ITD*) s'accroissant entre des signaux audio binauraux individuellement adaptables (L^* , R^*) est adaptée individuellement par la mise à l'échelle du facteur de conversion actuel (étape 130).
4. La méthode (100) de la revendication 3, **caractérisée en ce que** l'accroissement de la différence de temps de transit (ITD*) est adapté individuellement par la mise à l'échelle des différences de temps de transit (ITD) correspondant aux fonctions de transfert de tête ($xRIR^*$) dépourvues de temps de transit.
5. La méthode (100) de l'une des revendications précédentes, **caractérisée en ce que** les fonctions de transfert de tête ($xRIR^*$) dépourvues de temps de transit sont calculées pour une pluralité de positions de tête ($1...k$) et pour une pluralité de sources de signal audio ($1...n$) (étape 110).
6. La méthode (100) de l'une des revendications précédentes, **caractérisée en ce que** les temps de transit interauraux (TL, TR) sont extraits des fonctions de transfert de tête ($xRIR$) au moyens d'une des méthodes suivantes : méthode de détermination ; méthode de corrélation traverse interaurale ; par construction de différence (temps de transit) de délai de groupe spécifique de fréquences ; par correspondance de gradient de phase spécifique de gamme de fréquences ou par détermination de phase en excès au moyen d'une transformation Hilbert (étape 110).
7. La méthode (100) de la revendication 6, **caractérisée en ce que** la méthode de détermination est appliquée de telle façon qu'un plancher de bruit de mesure naturelle est déterminée et puis est déterminé un seuil situé au-dessus qui reste inférieur au maximum absolu de la réponse impulsionnelle résultant de la fonction de transfert de tête correspondante (étape 110).
8. La méthode (100) de la revendication 7, **caractérisée en ce que** le seuil est déterminé pour être au moins 10dB au dessus du plancher de bruit de mesure naturelle et/ou le seuil est déterminé pour être au moins 10dB plus bas que le maximum absolu de la réponse impulsionnelle résultant de la fonction de transfert de tête correspondante (étape 110).
9. La méthode (100) de la revendication 8, **caractérisée en ce que** le seuil est dans une gamme moyenne de la dynamique de mesure (étape 110).
10. La méthode (100) de la revendication 6, **caractérisée en ce que** la méthode de détermination est appliquée de manière à ce qu'une valeur en pourcentage du maximum absolu de la réponse impulsionnelle résultant d'une fonction de transfert de tête correspondante, est détectée pour être fixée, en particulier comme valeur entre 10% et 90% (étape 110).
11. Dispositif pour mettre en oeuvre une méthode suivant l'une des revendications précédentes, dans lequel le dispositif pour générer des signaux audio binauraux (L^* , R^*) pour une perception spatiale, comporte :

un premier circuit (A) pour extraire à partir des fonctions de transfert de tête ($xRIR$) qui sont associées à diverses positions de tête ($1...k$) des fonctions de transfert de tête ($xRiR^*$) dépourvues de temps de transit, et pour en calculer des différences de temps de transit (ITD) ;

un second circuit (BB) pour générer, à partir de signaux audio de synthèse binauraux (L' , R') dépourvus de temps de transit, chacun étant relatif à un chemin de signal binaural, dans lequel ce circuit utilise les fonctions de transfert de tête ($xRIR^*$), dépourvues de temps de transit, en fonction de données de position de tête (HTDAT) et ce pour une position de tête actuelle ($1...k$) ;

caractérisé par

au moins un troisième circuit (BA^* , C^*) pour générer, à partir des signaux audio (L' , R') dépourvus de temps de transit, les signaux audio binauraux individuellement adaptables (L^* , R^*) en chargeant pour au moins un chemin de signal le signal audio généré (L') dépourvu de temps de transit, avec un délai temporel (VDL) en fonction de la différence de temps de transit (ITD) calculé pour la position de tête courante ($1...k$) et en

EP 2 357 854 B1

fonction d'un facteur d'échelle individuel (ISF) ;
dans lequel le troisième circuit au moins (BA*; C*) change de manière adaptative le délai temporel (VDL) entre une première valeur et une seconde valeur au moyen d'une conversion du taux d'échantillonnage lorsque intervient un changement de la position de tête précédente (k=1) vers la position de tête actuelle (k=2), dans lequel la conversion de taux d'échantillonnage applique un facteur de conversion pour reproduire en avance ou en retard le signal audio dépourvu de temps de transit ; et
détermine le facteur de conversion utilisé pour la conversion du taux d'échantillonnage (SRC) en fonction du changement de délai temporel correspondant au changement de la position de tête précédente vers la position de tête actuelle.

5

10

15

20

25

30

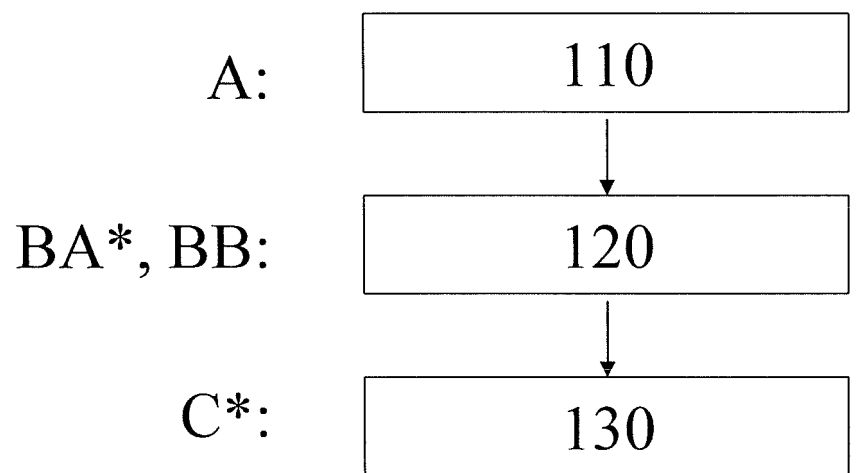
35

40

45

50

55



100

Fig. 1

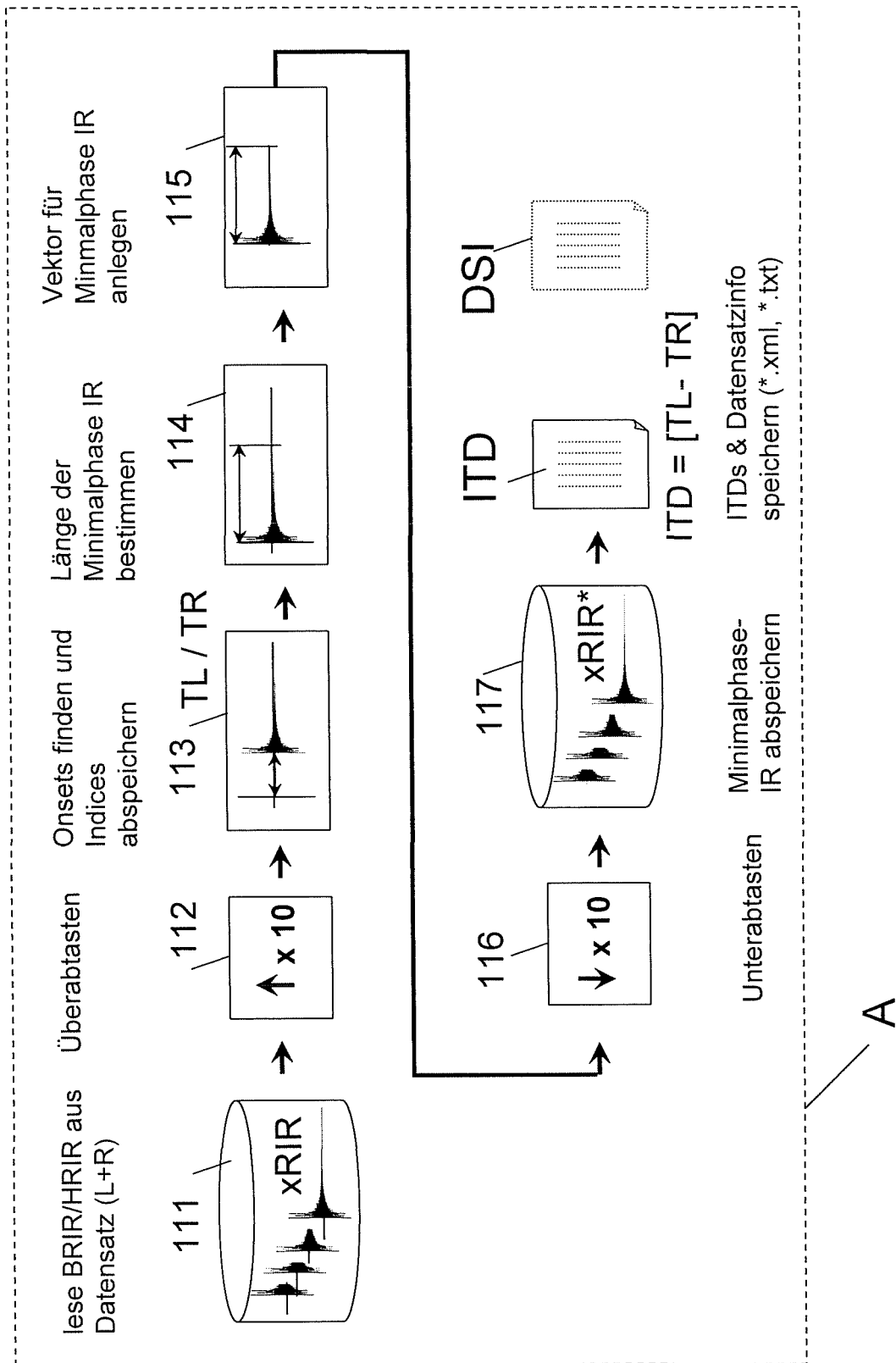


Fig. 2

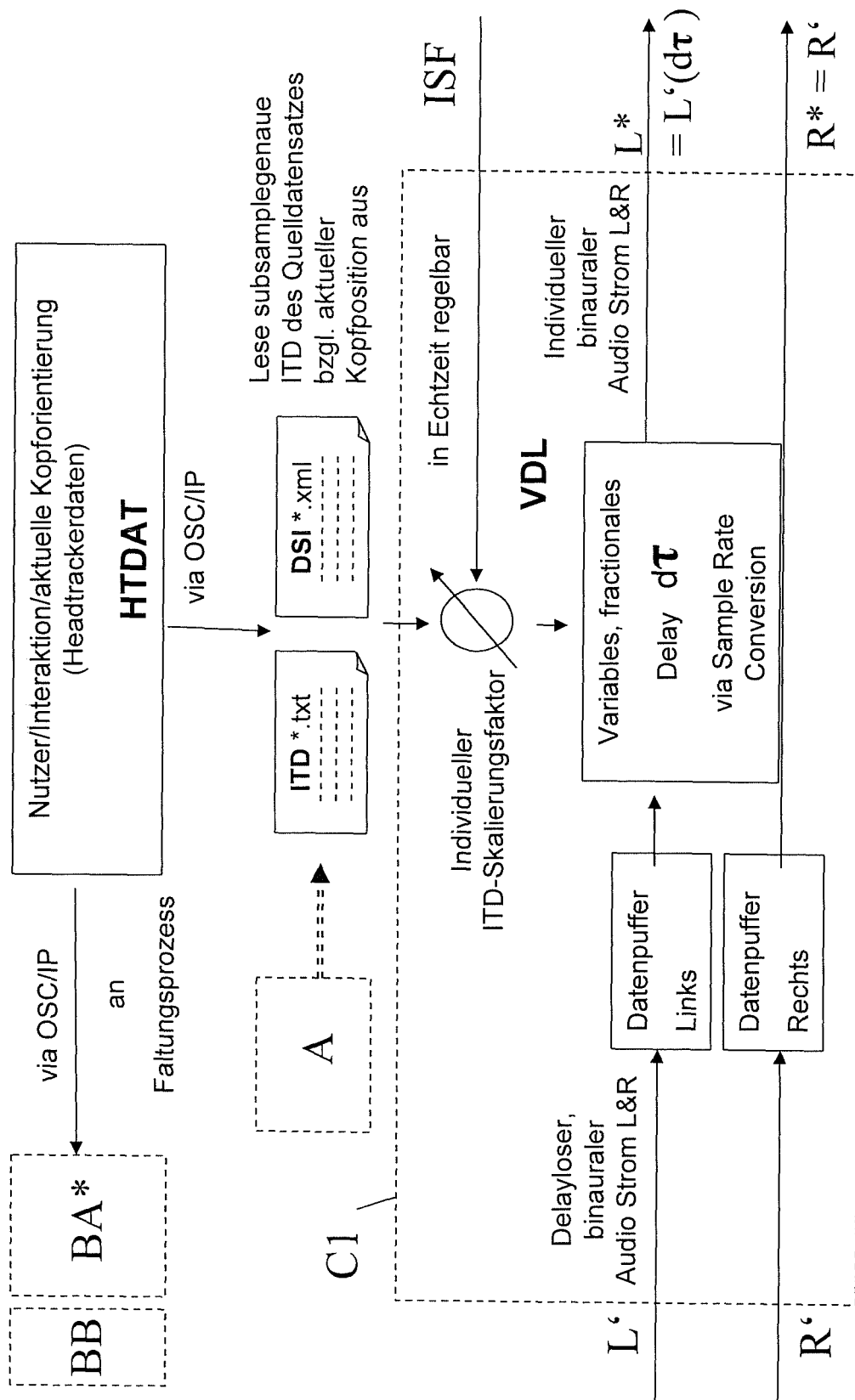


Fig. 3

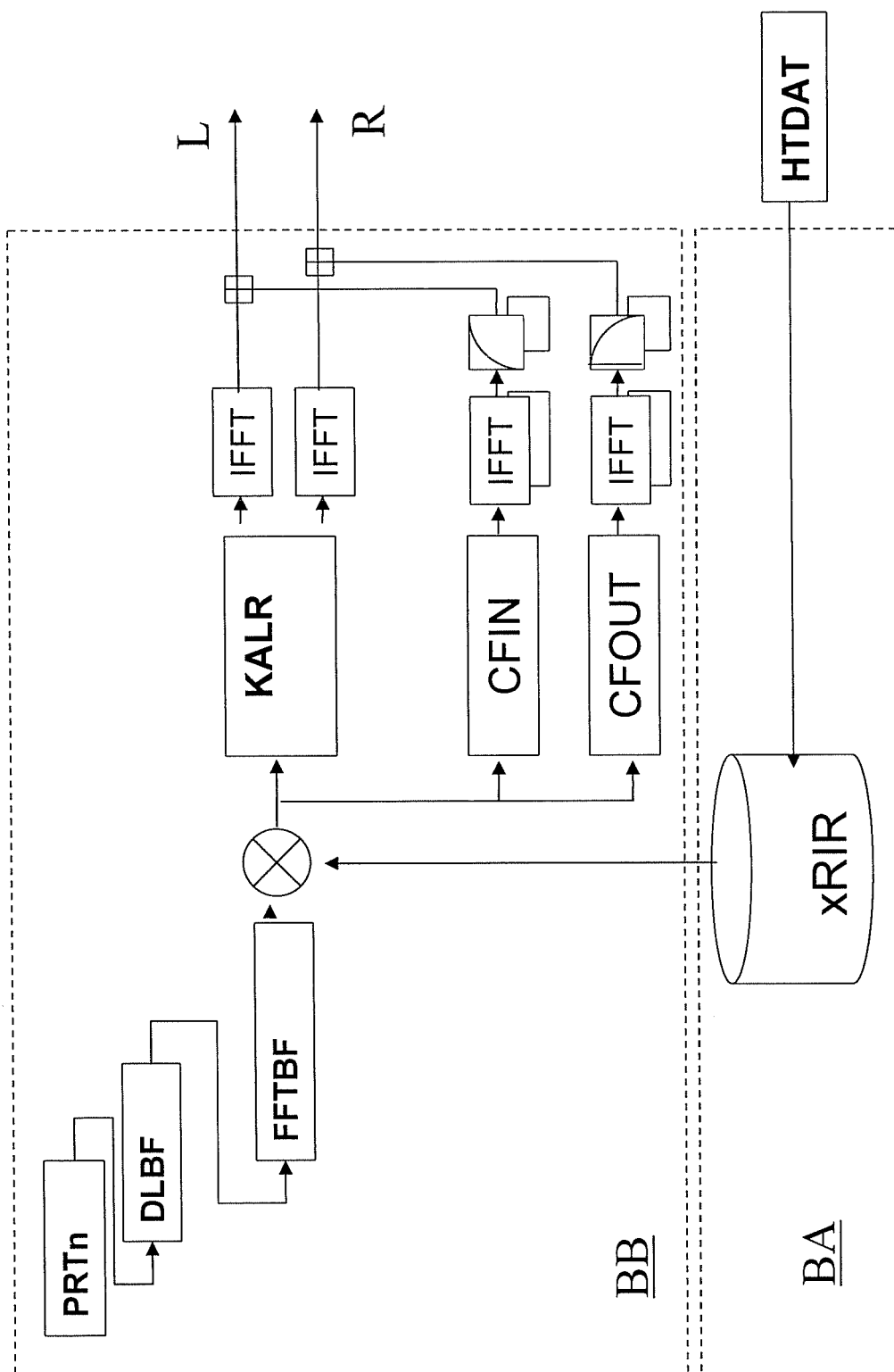


Fig. 4a

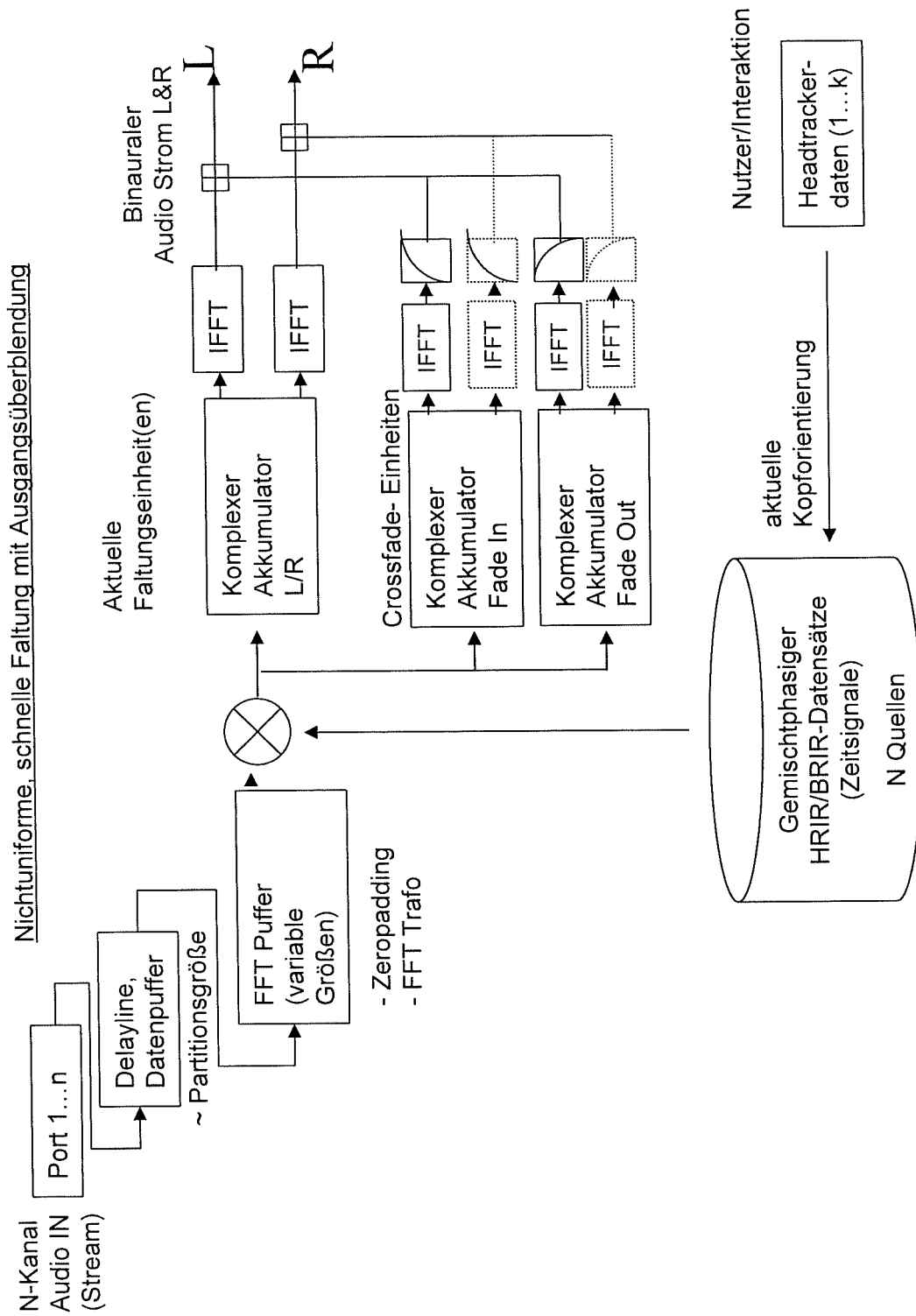


Fig. 4b

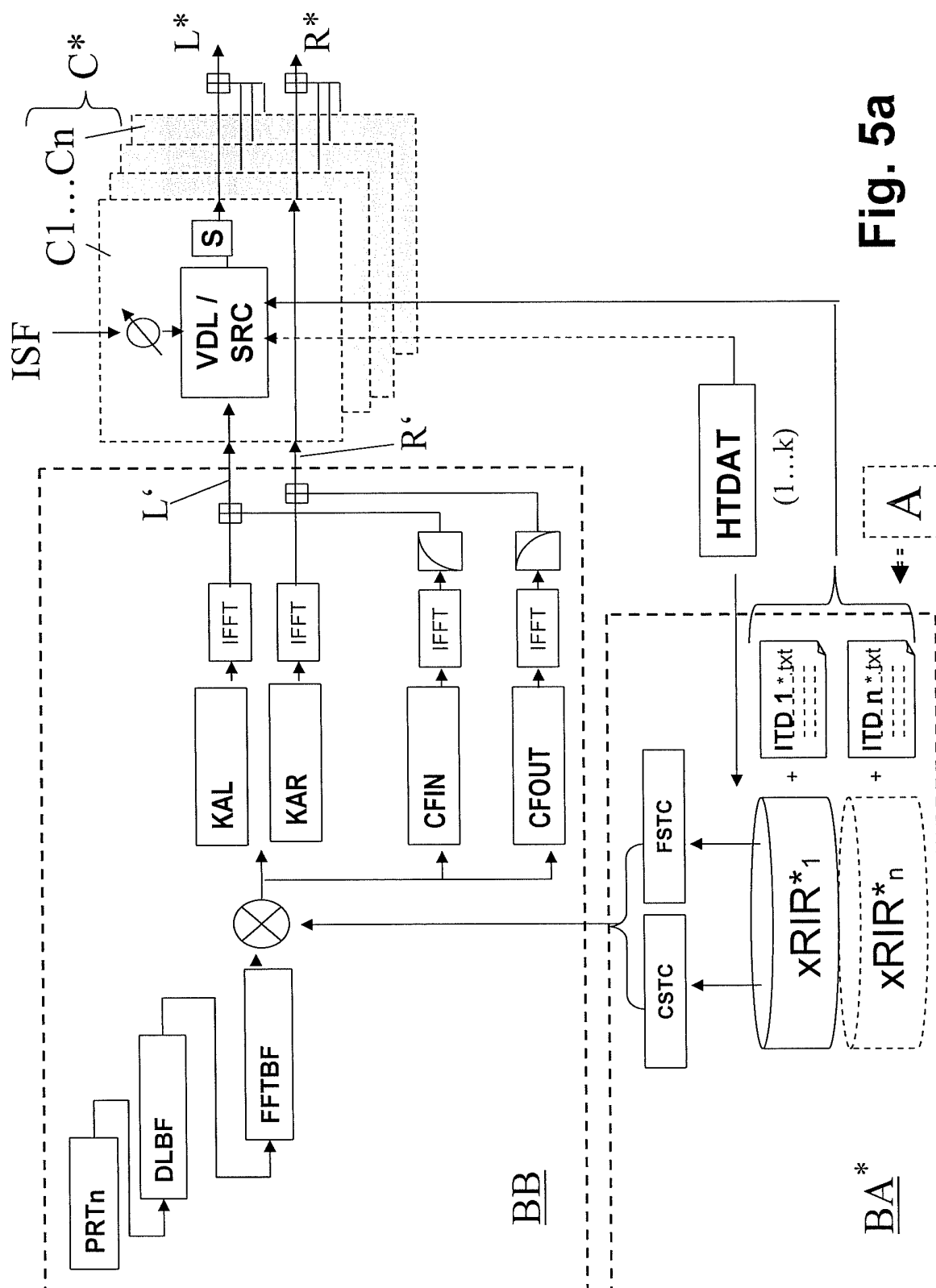


Fig. 5a

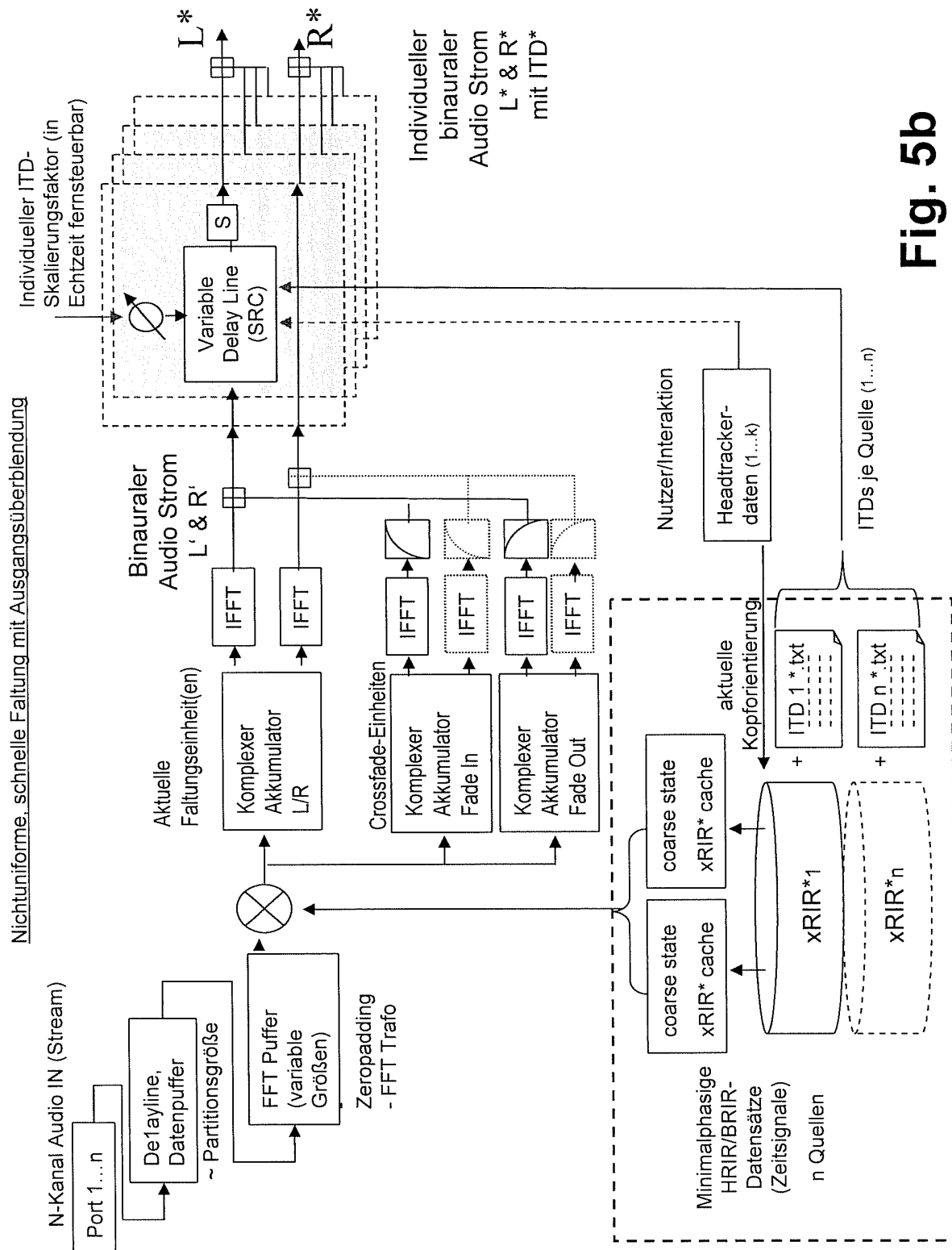


Fig. 5b

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

- GB 2369976 A [0004]
- US 7174229 B1 [0004] [0011]
- US 6795556 B [0004] [0012]
- EP 0912077 B1 [0004]
- DE 69523643 T2 [0004]
- US 2006045294 A1 [0009]

In der Beschreibung aufgeführte Nicht-Patentliteratur

- **LINDAU, A. ; WEINZIERL, S.** *Fabian - Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden*, 2007 [0007]
- **VON A. LINDAU ; S. WEINZIERL.** Fortschritte der Akustik. Tagungsband der 33. DAGA Stuttgart, Deutschland [0007]
- Finding the onset of a room impulse response: Straightforward?. **VON G. DEFRANCE et al.** The Journal of the Acoustical Society of America. Oktober 2008, vol. 124, 248-254 [0010]
- **SMITH III, J.O.** Digital Audio Resampling Home Page. Center for Computer Research in Music and Acoustics, 2002 [0054]