



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 60 2004 005 197 T2 2007.06.28**

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 636 791 B1**

(51) Int Cl.⁸: **G10L 19/12 (2006.01)**

(21) Deutsches Aktenzeichen: **60 2004 005 197.7**

(86) PCT-Aktenzeichen: **PCT/EP2004/006850**

(96) Europäisches Aktenzeichen: **04 740 263.1**

(87) PCT-Veröffentlichungs-Nr.: **WO 2005/001813**

(86) PCT-Anmeldetag: **24.06.2004**

(87) Veröffentlichungstag
der PCT-Anmeldung: **06.01.2005**

(97) Erstveröffentlichung durch das EPA: **22.03.2006**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **07.03.2007**

(47) Veröffentlichungstag im Patentblatt: **28.06.2007**

(30) Unionspriorität:
10328777 25.06.2003 DE

(84) Benannte Vertragsstaaten:
DE, FR, GB, IT

(73) Patentinhaber:
Coding Technologies AB, Stockholm, SE

(72) Erfinder:
**HÖRICH, Holger, 90419 Nürnberg, DE; SCHUG,
Michael, 91056 Erlangen, DE; NEUSINGER,
Matthias, 91189 Rohr, DE**

(74) Vertreter:
**Schoppe, Zimmermann, Stöckeler & Zinkler, 82049
Pullach**

(54) Bezeichnung: **VORRICHTUNG UND VERFAHREN ZUM KODIEREN EINES AUDIOSIGNALS UND VORRICHTUNG
UND VERFAHREN ZUM DEKODIEREN EINES KODIERTEN AUDIOSIGNALS**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf Codierungstechniken und insbesondere auf Audiocodierungstechniken. Audiocodierer und insbesondere solche Codierer, die unter dem Stichwort „mp3“, „AAC“ oder „mp3PRO“ bekannt sind, erlangten in letzter Zeit breite Akzeptanz. Sie ermöglichen die Kompression von Audiosignalen, die eine beträchtliche Datenmenge erfordern, wenn sie beispielsweise in einem PCM-Format auf einer Audio-CD vorliegen, zu „tolerierbaren“ Datenraten, die für die Übertragung der Audiosignale über Kanäle mit einer begrenzten Bandbreite hinweg geeignet sind. Zum Übertragen von Daten im PCM-Format sind somit Datenraten von bis zu 1,4 Mbit/s erforderlich. „mp3“-codierte Audiodaten erreichen bereits einen Stereoton mit hoher Qualität bei Datenraten von 128 kbit/s.

[0002] Ferner ist die Spektralbandvervielfältigung (SBR – spectral band replication) ein bekanntes Verfahren, das die Effizienz von existierenden an das Hörvermögen angepassten Wahrnehmungs-Audiocodierern bedeutend erhöht. Die SBR-Technik ist in der WO 98/57436 beschrieben und im „mp3PRO“-Format implementiert. Hier wird eine gute Stereoqualität bereits mit Datenraten von 64 kbit/s erzielt.

[0003] Die europäische Patentschrift EP 0 846 375 B1 offenbart ein Verfahren und eine Vorrichtung zur skalierbaren Codierung von Audiosignalen. Ein Audiosignal wird mittels eines ersten Codierers codiert, um den Bitstrom für den ersten Codierer zu erhalten. Dieses Signal wird anschließend erneut decodiert, mit einem Decodierer, der an den ersten Codierer angepasst ist. Das Decodier-Ausgangssignal wird zusammen mit dem verzögerten ursprünglichen Audiosignal an eine Differenzstufe geliefert, um ein Differenzsignal zu erzeugen. Dieses Differenzsignal wird Band für Band mit dem ursprünglichen Audiosignal verglichen, um für Spektralbänder zu ermitteln, ob die Energie des Differenzsignals größer als die Energie des Audiosignals ist. Falls dies der Fall ist, wird das ursprüngliche Audiosignal an einen zweiten Codierer geliefert, wohingegen das Differenzsignal an den zweiten Codierer geliefert wird, wenn die Energie des Differenzsignals geringer ist als die Energie des ursprünglichen Audiosignals. Der zweite Codierer ist ein Transformationscodierer, der auf der Basis eines psychoakustischen Modells fungiert. Wie der Bitstrom des ersten Codierers wird auch der Bitstrom auf der Ausgangsseite des zweiten Codierers einem Bitstrommultiplexer zugeführt, der auf der Ausgangsseite einen so genannten skalierten Bitstrom liefert. In diesem Zusammenhang bedeutet Skalierbarkeit, dass ein Decodierer je nach Entwurf in der Lage ist, entweder lediglich den Bitstrom des ersten Codierers aus dem Bitstrom auf der Decodierseite zu extrahieren, oder, sowohl den Bitstrom des ersten Codie-

ers als auch den Bitstrom des zweiten Codierers zu extrahieren, um im ersten Fall eine Reproduktion einer geringeren Qualität und im zweiten Fall eine qualitativ hochwertige Reproduktion des ursprünglichen Audiosignals zu erhalten.

[0004] Ein üblicherweise transformationsbasierter Codierer ist in [Fig. 4a](#) veranschaulicht. Das Audiosignal wird an eine Analysefilterbank **400** geliefert, die an ihrem Eingang einen Block mit einer bestimmten Anzahl von Abtastwerten des Audiosignals aus dem Strom von Abtastwerten mittels eines Blockierens bzw. Fensterns bildet und ihn in eine spektrale Darstellung umwandelt. Die spektralen Koeffizienten bzw. Teilbandsignale, die an dem Ausgang der Analysefilterbank erzeugt werden, werden quantisiert. Die Quantisiererschrittgröße hängt von verschiedenen Faktoren ab. Ein bedeutender Faktor ist eine psychoakustische Maskierungsschwelle, die durch ein psychoakustisches Modell **402** aus dem ursprünglichen Audiosignal berechnet wird. Der Quantisierer in einem Block „Quantisieren und Codieren **404**“ versucht immer, so grob wie möglich zu quantisieren, um eine gute Kompression zu erhalten. Dagegen versucht er jedoch auch, so fein wie möglich zu quantisieren, so dass das durch das Quantisieren bewirkte Quantisierungsrauschen unterhalb der durch den Block **402** gelieferten psychoakustischen Maskierungsschwelle liegt, was in der Technik hinreichend bekannt ist. Die auf diese Weise quantisierten Spektralwerte werden dann einer Entropiecodierung unterzogen, wobei üblicherweise eine Huffman-Codierung als Entropiecodierung verwendet wird, die üblicherweise mit vordefinierten Huffman-Codebüchern bzw. Huffman-Codetabellen arbeitet. Dann werden die entropiecodierten quantisierten Spektralwerte an den Ausgang des Blocks **404** angelegt, die zusammen mit den Nebeninformationen, die für das Decodieren mittels eines Blocks **406** benötigt werden, in einen Bitstrom **408** geschrieben werden, wobei dieser Bitstrom gespeichert oder, je nach dem Anwendungsgebiet, über einen Übertragungskanal hinweg an einen Decodierer übertragen werden kann, der in [Fig. 4b](#) veranschaulicht ist. Zunächst umfasst der Decodierer einen Block **410** zum Lesen des Bitstroms, um einerseits die Nebeninformationen und andererseits die entropiecodierten quantisierten Spektralwerte aus dem Bitstrom zu extrahieren. Dann werden die entropiecodierten quantisierten Spektralwerte zuerst einer Entropiedecodierung und anschließend einer Umkehr-Quantisierung zugeführt, um Umkehr-quantisierte Spektralwerte (Block **412**) zu erhalten, die anschließend über eine Synthesefilterbank **414**, die an die Analysefilterbank **400** angepasst ist, geliefert werden, um ein zeitdiskretes decodiertes Audiosignal auf der Ausgangsseite zu erhalten. Dieses zeitdiskrete Audiosignal an dem Ausgang der Synthesefilterbank kann anschließend nach einer entsprechenden Interpolation und Digital/Analog-Umwandlung und, falls notwendig, einer Verstär-

kung an einen Lautsprecher geliefert werden und dadurch hörbar gemacht werden. Blockbasierte Codierer/Decodierer, wie sie bei dem in [Fig. 4a](#) und [Fig. 4b](#) gezeigten bekannten Szenario verwendet werden, beruhen auf der Tatsache, dass üblicherweise ein Block von Abtastwerten, z. B. 1024 und 2048, mit einem in der Technik bekannten MDCT mit Überlappen- und Addieren-(Overlap and Add) zeitdiskreten Abtastwerten eines Audiosignals in den Spektralbereich umgewandelt werden. Sogar bei weniger stark frequenzauflösenden Filterbänken wie z. B. der SBR-Filterbank mit 64 Kanälen wird immer auch ein Block von Abtastwerten mit einer bestimmten Anzahl von Abtastwerten verwendet und in eine Spektraldarstellung umgewandelt, hier nämlich die einzelnen Teilbandsignale. Dann wird die spektrale Darstellung, wie erörtert wurde, entsprechend quantisiert, üblicherweise mit Hilfe eines psychoakustischen Modells, das die psychoakustische Maskierungsschwelle auf die in der Technik bekannte Weise berechnet.

[0005] Derartige Transformationen weisen inhärent eine bestimmte Zeit-/Frequenzauflösung auf. Dies bedeutet, dass, wenn eine große Anzahl von Abtastwerten in einen Block eingeführt wird, eine auf den Block angewendete Transformation inhärent eine hohe Frequenzauflösung aufweist. Dagegen wird die Zeitauflösung entsprechend verringert. Wenn die kürzeren Abschnitte des Audiosignals in den Spektralbereich umgewandelt würden, um die Zeitauflösung zu erhöhen, würde dies dazu führen, dass die Frequenzauflösung entsprechend Schaden nimmt.

[0006] Somit besteht ein Problem darin, dass Audiosignale nur für sehr kurze Zeiträume als ortsfest erachtet werden können. Es gibt gewiss kurzfristige starke Energieanstiege, die als Einschwingvorgang bzw. Transiente bezeichnet werden, während derer das Audiosignal nicht ortsfest ist.

[0007] Um dieses Problem der Zeit-/Frequenzauflösung anzugehen, wird beispielsweise bei dem AAC-Codierer (AAC = advanced audio coding, hochentwickeltes Audiocodieren) ein Blockschalten verwendet, das durch einen Transientendetektor gesteuert wird. Hier wird das zu codierende Audiosignal vor einem Fenstern bzw. Blockieren geprüft, um zu bestimmen, ob das Audiosignal eine derartige Transiente aufweist oder nicht. Falls eine Transiente ermittelt wird, werden zum Codieren kurze Blöcke verwendet. Wenn jedoch ein Signalabschnitt ohne Transiente erfasst wird, wird eine lange Blocklänge verwendet. Somit wird bei derartigen üblichen Transformationscodierungsverfahren ein Blockschalten verwendet, um die Transformationslänge an das Signal anzupassen. Besonders dann, wenn niedrige Bitraten erzielt werden sollen, werden vorzugsweise sehr lange Transformationslängen verwendet, da das Verhältnis von Seiteninformationen zu nützlichen Informationen üblicherweise relativ unabhängig von der

Blocklänge ist. Das bedeutet, dass die Menge an Seiteninformationen meistens dieselbe ist, unabhängig davon, ob der Block eine große Anzahl von Zeitabstastwerten des Audiosignals darstellt oder ob ein Block kurz ist, d. h. eine geringe Anzahl von Abtastwerten darstellt. Somit zielt man aus Gründen der Codierungseffizienz darauf ab, immer so lange Blocklängen wie möglich bzw. lange Transformationslängen in einem Transformationscodierer zu verwenden.

[0008] Dagegen muss für eine Transientenerfassung und ein Umschalten zu kurzen Fenstern bei dem Auftreten von nichtortsfesten Bereichen des Audiosignals eine Verarbeitungsanstrengung akzeptiert werden, die jedoch trotzdem dazu führt, dass das Signal in seiner codierten Form entweder nur mit einer guten Frequenzauflösung oder nur mit einer guten Zeitauflösung existiert.

[0009] Die Aufgabe der vorliegenden Erfindung besteht darin, ein verbessertes Konzept zum Codieren bzw. Decodieren zu liefern, um ein eine höhere Qualität aufweisendes und trotzdem effizientes Audiocodieren/-decodieren zu erhalten.

[0010] Diese Aufgabe wird durch eine Vorrichtung zum Codieren eines Audiosignals gemäß Anspruch 1, ein Verfahren zum Codieren eines Audiosignals gemäß Anspruch 7, eine Vorrichtung zum Decodieren eines codierten Audiosignals gemäß Anspruch 8, ein Verfahren zum Decodieren eines codierten Audiosignals gemäß Anspruch 9 oder ein Computerprogramm gemäß Anspruch 10 gelöst.

[0011] Die vorliegende Erfindung beruht auf dem Wissen, dass eine gute Codierungsqualität sowohl einer guten Frequenzauflösung als auch einer guten Zeitauflösung dadurch erzielt wird, dass im Sinn des Skalierbarkeitskonzepts ein erster Codierer eine erste Zeit-/Frequenzauflösung aufweist und dass ein zweiter Codierer eine zweite Zeit-/Frequenzauflösung aufweist, die sich voneinander unterscheiden, so dass der erste Codierer das ursprüngliche Audiosignal mit einer bestimmten Auflösung codiert, und dass der zweite Codierer dann mit einer bestimmten verschiedenen Auflösung bezüglich der Zeit bzw. der Frequenz arbeitet, so dass zwei Datenströme erhalten werden, die, wenn sie zusammen betrachtet werden, sowohl eine gute Zeitauflösung als auch eine gute Frequenzauflösung darstellen.

[0012] Außerdem wird nicht das ursprüngliche Audiosignal an den zweiten Codierer geliefert, sondern die Differenz zwischen dem ursprünglichen Audiosignal und dem codierten und wieder decodierten Ergebnis des ersten Codierers/Decodierers. Der Auflösungsfehler, den der erste Codierer machte, erscheint dann automatisch in dem Restsignal, das beispielsweise durch eine Differenzbildung erhalten

wird, wobei das Restsignal üblicherweise Fehler aufweist, beispielsweise aufgrund der schlechten Zeitauflösung des Erstcodier-/decodier-Pfades. Im Gegensatz dazu weist das Restsignal kaum diesbezügliche Frequenzfehler auf, da der Erstcodier-/decodier-Pfad eine gute Frequenzauflösung aufwies. Somit kann das Restsignal ohne weiteres mit einem Codierer mit einer hohen Zeitauflösung (und somit einer schlechten Frequenzauflösung) codiert werden, um ein Signal als Zweitcodierungsausgangssignal zu erhalten, das eine gute Zeitauflösung, jedoch eine schlechte Frequenzauflösung aufweist, was jedoch nichts ausmacht, da das Erstcodier-Ausgangssignal bereits eine gute Frequenzauflösung aufweist und somit die bezüglich der Frequenz betrachtete Struktur des Audiosignals sehr gut reproduziert.

[0013] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung sind sowohl der erste Codierer als auch der zweite Codierer Transformationscodierer. Ferner ist bevorzugt, dass der erste Codierer mit einer hohen Frequenzauflösung (und somit einer schlechten Zeitauflösung), d. h. mit einer hohen Transformationslänge, betrieben wird, wohingegen der zweite Codierer mit einer hohen Zeitauflösung (und somit einer schlechten Frequenzauflösung) betrieben wird.

[0014] Gemäß der Erfindung hat man festgestellt, dass Artefakte im Zeitbereich, was Artefakte bedeutet, die auf eine schlechte Zeitauflösung zurückzuführen sind, in vielen Fällen eher akzeptiert werden als Artefakte im Frequenzbereich, d. h. Artefakte, die auf eine schlechte Frequenzauflösung zurückzuführen sind. Somit ist bevorzugt, den ersten Codierer mit einer hohen Frequenzauflösung zu betreiben, da dann lediglich das Erstcodier-Ausgangssignal von einem jeweiligen Decodierer ausreichend ist, um eine einigermaßen gute Audioausgabe zu erhalten, was in dem Skalierbarkeitskonzept enthalten ist.

[0015] Gemäß der Erfindung wird die Qualität des Erstcodierverfahrens durch den zweiten Codierer verbessert, indem eine Differenzbildung zwischen dem Ausgangssignal des Erstcodier-/decodier-Pfades und dem ursprünglichen Audiosignal durchgeführt wird, und dass das resultierende Restsignal dann mit dem zweiten Codierer, der eine gute Zeitauflösung aufweist, codiert wird. Dieses Codieren ist besonders für das Restsignal günstig, da es bereits wenige tonale Elemente umfasst, da sie bereits durch das Erstcodierungsverfahren sehr gut und effizient festgehalten wurden.

[0016] Das bedeutende Defizit dieses Restsignals ist jedoch die schlechte Zeitauflösung, die sich in der Erzeugung eines Rauschens vor oder nach einer Transiente, d. h. eines Vorechos oder eines Nachechos, zeigt. Vorechos stören noch mehr als Nachechos, da sie für ein Subjekt ohne weiteres erfassbar

sind. Dieses Rauschen ist sozusagen das Quantisierungsrauschen der Transiente und entspricht in seinem Spektralgehalt hauptsächlich der einer der Transiente und ist somit nicht tonal. Somit wird die Zeitauflösung auf effiziente Weise beträchtlich verbessert, indem das Transformationscodierungsverfahren mit kürzeren Blöcken, d. h. mit einer hohen Zeitauflösung, verwendet wird.

[0017] Somit wird gemäß der Erfindung ein Audio-codierungsverfahren mit einer hohen und höchsten Qualität erhalten, indem die Abschnitte des Audiosignals, die nicht tonal oder eher tonal sind, mit einem frequenzselektiven Transformationscodierungsverfahren mit langen Transformationslängen erfasst werden, während ein in Verarbeitungsrichtung nachgelagertes Codierungsverfahren mit einer kurzen Transformationslänge eine hohe Zeitauflösung für das Restsignal ermöglicht.

[0018] Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend Bezug nehmend auf die beiliegenden Zeichnungen ausführlicher erörtert. Es zeigen:

[0019] [Fig. 1](#) ein Blockdiagramm eines erfindungsgemäßen Codierungskonzepts;

[0020] [Fig. 2](#) ein Blockdiagramm eines erfindungsgemäßen Codierungskonzepts gemäß einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung;

[0021] [Fig. 3](#) ein Blockdiagramm eines erfindungsgemäßen Decodiererkonzepts;

[0022] [Fig. 4a](#) einen bekannten Transformationscodierer; und

[0023] [Fig. 4b](#) einen bekannten Transformationsdecodierer.

[0024] [Fig. 1](#) zeigt eine Vorrichtung zum Codieren eines Audiosignals, das über einen Eingang **10** bereitgestellt wird. Zuerst wird das Audiosignal in einem ersten Codierer **12** mit einer ersten Zeit-/Frequenzauflösung zugeführt. Der erste Codierer **12** ist dahin gehend gebildet, ein Erstcodier-Ausgangssignal an einem Ausgang **14** zu erzeugen. Das Erstcodier-Ausgangssignal an dem Ausgang **14** des ersten Codierers **12** wird einerseits an einen Multiplexer **16** und andererseits an einen Decodierer **18** geliefert, der an den ersten Codierer angepasst ist und das Erstcodier-Ausgangssignal decodiert, um an einem Ausgang **20** des Decodierers **18** ein decodiertes Audiosignal zu liefern. Das decodierte Ausgangssignal **20** sowie das ursprüngliche Audiosignal **10** wird an eine Vergleichseinrichtung **22** geliefert. Die Vergleichseinrichtung **22** ist dahin gehend gebildet, das Audiosignal an dem Eingang **10** mit dem decodierten Audiosi-

gnal an dem Ausgang **20** zu vergleichen, was bedeutet nach dem Pfad von dem ersten Codierer **12** und dem Decodierer **18**. Die Vergleichseinrichtung **22** ist besonders dahin gehend geformt, ein Restsignal an einem ihrer Ausgänge **24** zu liefern, wobei das Restsignal eine Differenz zwischen dem Audiosignal und dem decodierten Audiosignal umfasst. Dieses Restsignal **24** wird an einen zweiten Codierer **26** geliefert, der dahin gehend gebildet ist, das Restsignal an dem Ausgang **24** der Vergleichseinrichtung **22** zu codieren, um ein Zweitcodier-Ausgangssignal an einem Ausgang **28** zu liefern, das auch an den Multiplexer **16** geliefert wird. Der Multiplexer **16** ist dahin gehend gebildet, das Erstcodier-Ausgangssignal und das Zweitcodier-Ausgangssignal zu kombinieren und daraus ein codiertes Audiosignal an einem Ausgang **30** zu erzeugen, falls nötig unter Berücksichtigung entsprechender Nebeninformationen und Bitstrom-syntax-Konventionen.

[0025] Gemäß der Erfindung weist der erste Codierer eine erste Zeit- oder Frequenzauflösung auf, und der zweite Codierer weist eine zweite Zeit- oder Frequenzauflösung auf. Gemäß der vorliegenden Erfindung unterscheiden sich die erste Auflösung des ersten Codierers und die zweite Auflösung des zweiten Codierers, so dass das Erstcodier-Ausgangssignal entweder bezüglich der Zeit oder der Frequenz gut codiert ist und dass das Zweitcodier-Ausgangssignal bezüglich der Frequenz oder der Zeit gut codiert ist, so dass das codierte Audiosignal an dem Ausgang des Multiplexers **16** sowohl eine hohe Zeitauflösung als auch eine hohe Frequenzauflösung aufweist.

[0026] Nachstehend wird unter Bezugnahme auf [Fig. 2](#) ein bevorzugtes Ausführungsbeispiel der vorliegenden Erfindung veranschaulicht. Hier wird ein Audiosignal **10** einer Verzögerung durch ein Verzögerungsbauglied **32** unterworfen, bevor es an die Vergleichseinrichtung **22** geliefert wird, die in [Fig. 2](#) als Differenzbauglied veranschaulicht ist, so dass bei dem in [Fig. 2](#) gezeigten bevorzugten Ausführungsbeispiel durch das Differenzbauglied **22** eine auf den Abtastwert bezogene Differenzbildung in Echtzeit zwischen dem decodierten Audiosignal an dem Ausgang des Decodierers **18** und dem (verzögerten) Audiosignal an dem Ausgang des Verzögerungsbauglieds **32** durchgeführt werden kann.

[0027] Bei dem in [Fig. 2](#) gezeigten Ausführungsbeispiel sind ferner der erste Codierer, d. h. der Codierer **12** in [Fig. 2](#), und der zweite Codierer **26**, der in [Fig. 2](#) als Differenzcodierer bezeichnet wird, dahin gehend gebildet, eine Transformationscodierung durchzuführen.

[0028] Ferner ist bevorzugt, dass der erste Codierer **12** eine Codierung mit einer langen Transformationslänge, d. h. einer hohen Frequenzauflösung und somit einer geringen Zeitauflösung, durchführt, wäh-

rend der zweite Codierer **26** eine Codierung mit einer kurzen Transformationslänge, was eine hohe Zeitauflösung und einhergehend damit eine geringe Frequenzauflösung bedeutet, durchführt.

[0029] Obwohl der erste Codierer im Prinzip auch mit kurzen Transformationslängen und der Differenzcodierer mit langen Transformationslängen arbeiten könnten, ist es trotzdem bevorzugt, den ersten Codierer mit langen Transformationslängen zu betreiben, da, wie bereits erläutert wurde, Zeitartefakte eher weniger problematisch für einen Zuhörer sind als Frequenzartefakte. Somit kann ein Codierer, der lediglich das Erstcodier-Ausgangssignal an dem Ausgang **14**, jedoch nicht das Zweitcodier-Ausgangssignal an dem Ausgang **28** verarbeiten kann, eine angenehmere Reproduktion erzeugen, wenn der erste Codierer mit langen Transformationslängen arbeitet als wenn der erste Codierer mit kurzen Transformationslängen arbeiten würde.

[0030] Jegliche Einrichtung zum Umwandeln eines Blocks von Zeitabstastwerten in eine spektrale Darstellung kann als Transformationsalgorithmus in dem ersten Codierer und/oder dem zweiten Codierer der [Fig. 2](#) verwendet werden, z. B. eine Fourier-Transformation, eine diskrete Fourier-Transformation, eine schnelle Fourier-Transformation, eine diskrete Cosinus-Transformation, eine modifizierte diskrete Cosinus-Transformation usw. Alternativ dazu kann eine Filterbank mit einer geringen Anzahl von Kanälen verwendet werden, z. B. eine 64-Kanal-Filterbank, eine 128-Kanal-Filterbank oder eine Filterbank mit mehr oder weniger Kanälen.

[0031] Bei einem anderen Ausführungsbeispiel der vorliegenden Erfindung kann der erste Codierer **12** ein SBR-Codierer sein, der dahin gehend gebildet ist, ein Erstcodier-Ausgangssignal zu liefern, das lediglich Informationen bis zu einer Grenzfrequenz umfasst, die geringer ist als die Grenzfrequenz des Audiosignals an dem Audioeingang **10**. Typische SBR-Codierer extrahieren Nebeninformationen aus dem Audiosignal, die für eine Hochfrequenz-Rekonstruktion in einem SBR-Decodierer verwendet werden können, um das hohe Band, d. h. das Band des Audiosignals, das über der Grenzfrequenz des Erstcodier-Ausgangssignals liegt, mit einer so hohen Qualität wie möglich zu rekonstruieren. Jedoch ist der Decodierer **18** in [Fig. 2](#) kein derartiger SBR-Decodierer mit einer Hochfrequenz-Rekonstruktion, sondern ein üblicher Transformationsdecodierer, der an den ersten Codierer **12** angepasst ist, um das Codierausgangssignal unabhängig davon, dass dasselbe Band begrenzt ist, einfach zu decodieren, so dass das Ausgangssignal des Decodierers **18** an dem Ausgang **20** ebenfalls eine geringere Grenzfrequenz aufweist als das ursprüngliche Audiosignal.

[0032] In diesem Fall würde das Restsignal bis zu

der Grenzfrequenz den Codier-/Decodierfehler des Pfades des Codierers **12** und des Decodierers umfassen, wäre jedoch das vollständige Audiosignal oberhalb der Grenzfrequenz.

[0033] In diesem Fall könnte das Restsignal entweder auch mit einem Differenzcodierer **26** codiert werden, der kurze Transformationslängen verwendet, da es dem ursprünglichen Audiosignal oberhalb der Grenzfrequenz des Erstcodier-Ausgangssignals entspricht. Alternativ könnte jedoch lediglich der Spektralbereich des Restsignals bis zu der Grenzfrequenz des Erstcodier-Ausgangssignals mit dem Differenzcodierer **26** codiert werden, während der hochfrequente Abschnitt des Restsignals erneut mit dem ersten Codierer **12** mit den langen Transformationslängen codiert wird, um auch eine Hochfrequenzauflösung in dem hochfrequenten Teil des Audiosignals zu erhalten.

[0034] Das Ausgangssignal des Codierers **12** für das Hochfrequenzband kann anschließend erneut mit dem jeweiligen Band des ursprünglichen Audiosignals verglichen werden, um das Differenzsignal erneut mit dem Differenzcodierer **26** zu codieren, so dass am Ende vier Datenströme an den Multiplexer **16** geliefert werden, die, wenn sie alle miteinander decodiert werden, eine transparente Reproduktion, d. h. eine Reproduktion ohne Artefakte, ermöglichen.

[0035] Gemäß der Erfindung ist es nicht von Bedeutung, dass der erste Codierer und der zweite Codierer unter Verwendung eines psychoakustischen Modells arbeiten. Jedoch ist es aus Gründen der Dateneffizienz bevorzugt, dass zumindest der erste Codierer **12** unter Verwendung eines psychoakustischen Modells arbeitet. In Abhängigkeit von den Ressourcen könnte der zweite Codierer dann auf verlustfreie Weise codieren, wenn jeweilige Übertragungskanalressourcen vorliegen, so dass eine vollständig transparente Reproduktion erzielt wird. Alternativ dazu könnte der zweite Codierer ebenfalls unter Verwendung eines psychoakustischen Modells arbeiten, wobei bevorzugt ist, dass das psychoakustische Modell in diesem Fall nicht erneut vollständig für den zweiten Codierer berechnet wird, sondern dass zumindest Teile desselben bzw. die ganze psychoakustische Maskierungsschwelle unter Berücksichtigung der unterschiedlichen Transformationslängen des ersten Codierers im Vergleich zu dem zweiten Codierer „wieder verwendet“ werden kann bzw. können. Dies kann beispielsweise dadurch erfolgen, dass die psychoakustische Maskierungsschwelle, die durch den ersten Codierer berechnet wird, unmittelbar für den zweiten Codierer verwendet wird, wobei jedoch beispielsweise eine „Sicherheits-Überlast“ von z. B. 3 dB verwendet wird, um die kürzeren Transformationslängen des zweiten Codierers zu berücksichtigen, so dass die psychoakustische Maskierungsschwelle für den zweiten Codierer um z. B. 3 dB oder

einen anderen vorbestimmten Betrag geringer ist als die psychoakustische Maskierungsschwelle für den ersten Codierer **12**.

[0036] Bezüglich der Transformationslängen ist bevorzugt, dass die Transformationslänge des ersten Codierers eine ganzzahlige Vielzahl der Transformationslänge des zweiten Codierers ist. Auf diese Weise kann die Transformationslänge des ersten Codierers beispielsweise doppelt so viele, dreimal so viele, viermal so viele oder fünfmal so viele Abtastwerte des Audiosignals aufweisen wie die Transformationslänge des zweiten Codierers **26**. Diese ganzzahlige Beziehung zwischen der Transformationslänge des ersten und des zweiten Codierers ist deshalb bevorzugt, da eine relativ gute Wiederverwendung von Codierdaten des ersten Codierers für den zweiten Codierer möglich wird. Dagegen wäre auch eine nicht-ganzzahlige Beziehung zwischen der Transformationslänge unproblematisch, da der erste Codierer **12** und der zweite Codierer **26** auch auf eine nicht aufeinander synchronisierte Weise laufen können, solange dies entsprechend einem Decodierer berichtet wird, so dass derselbe die Summierung mit den korrekten Abtastwerten durchführt, was die Umkehrung der auf Abtastwerte bezogenen Differenzbildung bei dem Element **22** der [Fig. 2](#) bedeutet.

[0037] [Fig. 3](#) zeigt einen Decodierer zum Decodieren eines codierten Audiosignals gemäß der vorliegenden Erfindung. Das codierte Audiosignal, das an dem Ausgang **30** der [Fig. 1](#) bzw. der [Fig. 2](#) ausgegeben wird, wird nach einer Übertragung, Speicherung usw. an einen Eingang **40** des Decodierers in [Fig. 3](#) geliefert. Der Eingang **40** ist zuerst mit einer Extraktionseinrichtung **42** gekoppelt, die die Funktionalität eines Bitstromdemultiplexers aufweist, um zuerst das Erstcodier-Ausgangssignal aus dem codierten Audiosignal zu extrahieren und es an einem Ausgang **44** bereitzustellen, und die ferner dahin gehend gebildet ist, das codierte Restsignal bzw. das Differenzsignal bzw. das zweite codierte Audiosignal an einem Ausgang **46** bereitzustellen. Das Erstcodier-Ausgangssignal wird an einen ersten Decodierer geliefert, der an den ersten Codierer **12** der erfindungsgemäßen Vorrichtung zum Codieren, die in [Fig. 1](#) gezeigt ist, angepasst ist und im Prinzip identisch mit dem Decodierer **18** der [Fig. 1](#) sein kann. Dies bedeutet, dass der erste Decodierer **48** erneut dieselbe Zeit-/Frequenzauflösung aufweist, was bedeutet, dass er beispielsweise mit derselben Transformationslänge arbeitet wie der Codierer **12** der [Fig. 1](#). Das Zweitcodier-Ausgangssignal an dem Ausgang **46** der Extraktionseinrichtung wird an einen zweiten Decodierer **50** geliefert, der an den zweiten Codierer **26** der [Fig. 1](#) angepasst ist und somit die zweite Zeit-/Frequenzauflösung aufweist, was eine Zeit-/Frequenzauflösung bedeutet, die identisch mit der Zeit-/Frequenzauflösung des zweiten Codierers **26** in [Fig. 1](#) ist.

[0038] Auf der Ausgangsseite stellt der erste Codierer **48** das decodierte Audiosignal, das identisch mit dem Signal an dem Ausgang **20** der [Fig. 2](#) sein kann, bereit. Analog dazu stellt der zweite Decodierer **50** das decodierte Restsignal an seinem Ausgang bereit. Es ist zu beachten, dass beide Decodierer im Prinzip gemäß der Veranschaulichung in Bezug auf [Fig. 4b](#) gebildet sein können, wobei sich dieselben jedoch bezüglich ihrer Transformationslängen und somit bezüglich der verwendeten Synthesefilterbänke unterscheiden können.

[0039] Sowohl das decodierte Audiosignal an dem Ausgang **52** in [Fig. 3](#) als auch das decodierte Restsignal an dem Ausgang **54** der [Fig. 3](#) werden an eine Kombinationseinrichtung **56** geliefert, die bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung eine auf Abtastwerte bezogene Summierung durchführt, was allgemein eine Operation bedeutet, die umgekehrt zu der Vergleichsoperation ist, die bei dem Codierer in dem Element **22** der [Fig. 1](#) durchgeführt wurde. Auf der Ausgangsseite stellt die Kombinationseinrichtung **56** an einem Ausgang **58** der Decodiervorrichtung der [Fig. 3](#) ein Ausgangssignal bereit, das aufgrund der vorliegenden Erfindung sowohl durch eine gute Zeitauflösung als auch eine gute Frequenzauflösung herausragt, d. h. es umfasst sowohl wenige Frequenzartefakte als auch wenige Zeitartefakte.

[0040] Je nach den Umständen kann das erfindungsgemäße Verfahren zum Codieren, wie es unter Bezug auf [Fig. 1](#) veranschaulicht ist, oder das erfindungsgemäße Verfahren zum Decodieren, wie es unter Bezug auf [Fig. 3](#) veranschaulicht ist, in Hardware oder in Software implementiert sein. Die Implementierung kann an einem digitalen Speichermedium, besonders einer Platte oder einer CD mit elektronisch lesbaren Steuersignalen, durchgeführt werden, die mit einem programmierbaren Computersystem interagieren kann, so dass das jeweilige Verfahren ausgeführt wird. Somit besteht die Erfindung allgemein auch aus einem Computerprogrammprodukt mit einem Programmcode, der auf einem maschinenlesbaren Träger gespeichert ist, zum Durchführen des erfindungsgemäßen Verfahrens, wenn das Computerprogrammprodukt auf einem Computer läuft. Mit anderen Worten kann die Erfindung auch als Computerprogramm mit einem Programmcode zum Durchführen des Verfahrens verwirklicht werden, wenn das Computerprogramm auf einem Computer läuft.

Patentansprüche

1. Vorrichtung zum Codieren eines Audiosignals, die folgende Merkmale aufweist:
einen ersten Transformationscodierer (**12**) zum Erzeugen eines Erstcodier-Ausgangssignals aus dem Audiosignal, wobei der erste Transformationscodie-

rer dahin gehend angepasst ist, einen Block mit einer ersten Anzahl von Zeitabtastwerten des Audiosignals in eine spektrale Darstellung umzuwandeln, um das Erstcodier-Ausgangssignal zu erhalten;
einen Decodierer (**18**), der an den ersten Codierer (**12**) angepasst ist, zum Decodieren des Erstcodier-Ausgangssignals, um ein decodiertes Audiosignal bereitzustellen;
eine Vergleichseinrichtung (**22**) zum Vergleichen des Audiosignals mit dem decodierten Audiosignal, wobei die Vergleichseinrichtung (**22**) dahin gehend angepasst ist, ein Restsignal bereitzustellen, wobei das Restsignal eine Differenz zwischen dem Audiosignal und dem decodierten Audiosignal umfasst;
einen zweiten Transformationscodierer (**26**) zum Codieren des Restsignals, um ein Zweitcodier-Ausgangssignal bereitzustellen, wobei der zweite Transformationscodierer dahin gehend angepasst ist, einen Block mit einer zweiten Anzahl von Zeitabtastwerten des Audiosignals in eine spektrale Darstellung umzuwandeln, um das Zweitcodier-Ausgangssignal zu erhalten,
wobei der erste Transformationscodierer und der zweite Transformationscodierer angepasst sind, so dass die erste Anzahl von Zeitabtastwerten des Audiosignals größer ist als die zweite Anzahl von Zeitabtastwerten des Audiosignals, und dass der erste Codierer (**12**) eine geringe Zeitauflösung und eine hohe Frequenzauflösung aufweist und dass der zweite Codierer (**26**) eine hohe Zeitauflösung und eine geringe Frequenzauflösung aufweist; und
einen Multiplexer (**16**) zum Kombinieren des Erstcodier-Ausgangssignals und des Zweitcodier-Ausgangssignals, um ein codiertes Audiosignal zu erhalten.

2. Vorrichtung gemäß Anspruch 1, bei der der erste Codierer (**12**) und der zweite Codierer (**26**) eine Filterbank oder einen Transformationsalgorithmus aufweisen, die beziehungsweise der eine Fourier-Transformation, eine diskrete Fourier-Transformation, eine schnelle Fourier-Transformation, eine diskrete Cosinus-Transformation oder eine modifizierte diskrete Cosinus-Transformation umfasst.

3. Vorrichtung gemäß Anspruch 1 oder Anspruch 2, bei der der Decodierer (**18**) dahin gehend angepasst ist, ein zeitdiskretes decodiertes Audiosignal mit einer Sequenz von Abtastwerten bereitzustellen, wobei das Audiosignal ein zeitdiskretes Audiosignal mit einer Sequenz von Abtastwerten ist, und wobei die Vergleichseinrichtung (**22**) dahin gehend ausgelegt ist, eine auf Abtastwerte bezogene Differenzbildung durchzuführen, um das Restsignal zu erhalten.

4. Vorrichtung gemäß einem der vorhergehenden Ansprüche, die ferner folgendes Merkmal aufweist:
ein Verzögerungsbauglied (**32**) zum Verzögern des

Audiosignals, wobei das Verzögerungsbauglied (**32**) dahin gehend angepasst ist, eine Verzögerung aufzuweisen, die von einer Verzögerung abhängt, die dem ersten Codierer (**12**) und dem Decodierer (**18**) zugeordnet ist.

5. Vorrichtung gemäß einem der vorhergehenden Ansprüche, bei der der Multiplexer (**16**) dahin gehend angepasst ist, das codierte Audiosignal zu erzeugen, so dass das Erstcodier-Ausgangssignal unabhängig von dem Zweitcodier-Ausgangssignal decodiert werden kann.

6. Vorrichtung gemäß einem der vorhergehenden Ansprüche, bei der der erste Codierer (**12**) dahin gehend angepasst ist, das Audiosignal einer Bandbegrenzung zu unterwerfen, so dass das Erstcodier-Ausgangssignal eine obere Grenzfrequenz aufweist, die geringer ist als eine obere Grenzfrequenz des Audiosignals, wobei die Vergleichseinrichtung (**22**) ein Restsignal liefert, das dem Audiosignal oberhalb der oberen Grenzfrequenz des Erstcodier-Ausgangssignals entspricht, und wobei der zweite Codierer (**26**) dahin gehend angepasst ist, einen Abschnitt des Restsignals oberhalb der oberen Grenzfrequenz des ersten Codierers mit einer Zeit- oder Frequenzauflösung zu codieren, die sich von der zweiten Auflösung unterscheidet oder gleich der zweiten Auflösung ist.

7. Verfahren zum Codieren eines Audiosignals, das folgende Schritte umfasst:
 Erzeugen (**12**) eines ersten Ausgangssignals mit einer ersten Zeit- oder Frequenzauflösung aus dem Audiosignal, wobei der Schritt des Erzeugens (**12**) den Schritt des Umwandeln eines Blocks mit einer ersten Anzahl von Zeitabtastrwerten des Audiosignals in eine spektrale Darstellung, um das erste Ausgangssignal zu erhalten, umfasst;
 Decodieren des Erstcodier-Ausgangssignals, um ein decodiertes Audiosignal bereitzustellen;
 Vergleichen (**22**) des Audiosignals mit dem decodierten Audiosignal, um ein Restsignal bereitzustellen, wobei das Restsignal eine Differenz zwischen dem Audiosignal und den decodierten Audiosignalen umfasst;
 Codieren (**26**) des Restsignals mit einer zweiten Zeit- oder Frequenzauflösung, um ein zweites Ausgangssignal zu liefern, wobei der Schritt des Codierens (**26**) den Schritt des Umwandeln eines Blocks mit einer zweiten Anzahl von Zeitabtastrwerten des Audiosignals in eine spektrale Darstellung, um das zweite Ausgangssignal zu erhalten, umfasst;
 wobei der Schritt des Erzeugens (**12**) und der Schritt des Codierens (**26**) dahin gehend angepasst sind, dass die erste Anzahl von Zeitabtastrwerten des Audiosignals größer ist als die zweite Anzahl von Zeitabtastrwerten des Audiosignals, und dass das erste Ausgangssignal eine geringe Zeitauflösung und eine hohe Frequenzauflösung aufweist und dass das

zweite Ausgangssignal eine hohe Zeitauflösung und eine geringe Frequenzauflösung aufweist; und
 Kombinieren (**16**) des Erstcodier-Ausgangssignals und des Zweitcodier-Ausgangssignals, um ein codiertes Audiosignal zu erhalten.

8. Vorrichtung zum Decodieren eines codierten Audiosignals, um ein Ausgangssignal zu erhalten, wobei das codierte Audiosignal ein Erstcodier-Ausgangssignal aufweist, das mit einer geringen Zeitauflösung und einer hohen Frequenzauflösung codiert ist, und wobei das codierte Audiosignal ferner ein Zweitcodier-Ausgangssignal aufweist, das ein Restsignal darstellt, das mit einer hohen Zeitauflösung und einer geringen Frequenzauflösung codiert ist, was einen Unterschied zwischen einem ursprünglichen Audiosignal und einem decodierten Audiosignal darstellt, wobei das decodierte Audiosignal durch ein Decodieren des Erstcodier-Ausgangssignals erhalten wird, wobei das Erstcodier-Ausgangssignal unter Verwendung eines ersten Transformationscodierers erzeugt wurde, wobei der erste Transformationscodierer dahin gehend angepasst ist, einen Block mit einer hohen Anzahl von Zeitabtastrwerten des Audiosignals in eine spektrale Darstellung umzuwandeln, um das Erstcodier-Ausgangssignal zu erhalten, wobei das Zweitcodier-Ausgangssignals unter Verwendung eines zweiten Transformationscodierers erzeugt wurde, und wobei der zweite Transformationscodierer dahin gehend angepasst ist, einen Block mit einer geringen Anzahl von Zeitabtastrwerten des Audiosignals in eine spektrale Darstellung umzuwandeln, um das Zweitcodier-Ausgangssignal zu erhalten, wobei die Vorrichtung folgende Merkmale aufweist:

eine Extraktionseinrichtung (**42**) zum Extrahieren des Erstcodier-Ausgangssignals und des Zweitcodier-Ausgangssignals aus dem codierten Audiosignal;
 einen ersten Transformationsdecoder (**48**), der an den ersten Transformationscodierer angepasst ist, zum Decodieren des Erstcodier-Ausgangssignals, um das decodierte Audiosignal zu erhalten, wobei der erste Decodierer (**48**) dahin gehend angepasst ist, mit der geringen Zeitauflösung und der hohen Frequenzauflösung zu arbeiten, und wobei der erste Transformationsdecoder (**48**) dahin gehend angepasst ist, einen Block mit einer ersten Anzahl von Spektralwerten in eine zeitliche Darstellung umzuwandeln;
 einen zweiten Transformationsdecoder (**50**), der an den zweiten Transformationscodierer angepasst ist, zum Decodieren des Zweitcodier-Ausgangssignals, um ein decodiertes Restsignal zu erhalten, wobei der zweite Decodierer dahin gehend angepasst ist, mit der hohen Zeitauflösung und der geringen Frequenzauflösung zu arbeiten, und wobei der zweite Transformationsdecoder (**50**) dahin gehend angepasst ist, einen Block mit einer zweiten Anzahl von Spektralwerten in eine zeitliche Darstellung umzu-

wandeln, wobei die zweite Anzahl geringer ist als die erste Anzahl, und eine Kombinationseinrichtung (56) zum Kombinieren des decodierten Audiosignals und des decodierten Restsignals, um das Ausgangssignal zu erhalten.

9. Verfahren zum Decodieren eines codierten Audiosignals, um ein Ausgangssignal zu erhalten, wobei das codierte Audiosignal ein Erstcodier-Ausgangssignal aufweist, das mit einer geringen Zeitauflösung und einer hohen Frequenzauflösung codiert ist, und wobei das codierte Audiosignal ferner ein Zweitcodier-Ausgangssignal aufweist, das ein Restsignal darstellt, das mit einer hohen Zeitauflösung und einer geringen Frequenzauflösung codiert ist, was einen Unterschied zwischen einem ursprünglichen Audiosignal und einem decodierten Audiosignal darstellt, wobei das decodierte Audiosignal durch ein Decodieren des Erstcodier-Ausgangssignals erhalten wird, wobei das Erstcodier-Ausgangssignal unter Verwendung eines ersten Transformationscodierers erzeugt wurde, wobei der erste Transformationscodierer dahin gehend angepasst ist, einen Block mit einer hohen Anzahl von Zeitabstastwerten des Audiosignals in eine spektrale Darstellung umzuwandeln, um das Erstcodier-Ausgangssignal zu erhalten, wobei das Zweitcodier-Ausgangssignal unter Verwendung eines zweiten Transformationscodierers erzeugt wurde, und wobei der zweite Transformationscodierer dahin gehend angepasst ist, einen Block mit einer geringen Anzahl von Zeitabstastwerten des Audiosignals in eine spektrale Darstellung umzuwandeln, um das Zweitcodier-Ausgangssignal zu erhalten, wobei das Verfahren folgende Schritte umfasst: Extrahieren (42) des Erstcodier-Ausgangssignals und des Zweitcodier-Ausgangssignals aus dem codierten Audiosignal;

Decodieren (48), angepasst an den ersten Transformationscodierer, des Erstcodier-Ausgangssignals, um das decodierte Audiosignal zu erhalten, wobei der Schritt des Decodierens (48) dahin gehend angepasst ist, mit der geringen Zeitauflösung und der hohen Frequenzauflösung zu arbeiten, und wobei der Schritt des Decodierens (98) dahin gehend angepasst ist, einen Block mit einer ersten Anzahl von Spektralwerten in eine zeitliche Darstellung umzuwandeln;

Decodieren (50), angepasst an den zweiten Transformationscodierer, des Zweitcodier-Ausgangssignals, um ein decodiertes Restsignal zu erhalten, wobei der Schritt des Decodierens dahin gehend angepasst ist, mit der hohen Zeitauflösung und der geringen Frequenzauflösung zu arbeiten, und wobei der Schritt des Decodierens (50) dahin gehend angepasst ist, einen Block mit einer zweiten Anzahl von Spektralwerten in eine zeitliche Darstellung umzuwandeln, wobei die zweite Anzahl geringer ist als die erste Anzahl, und

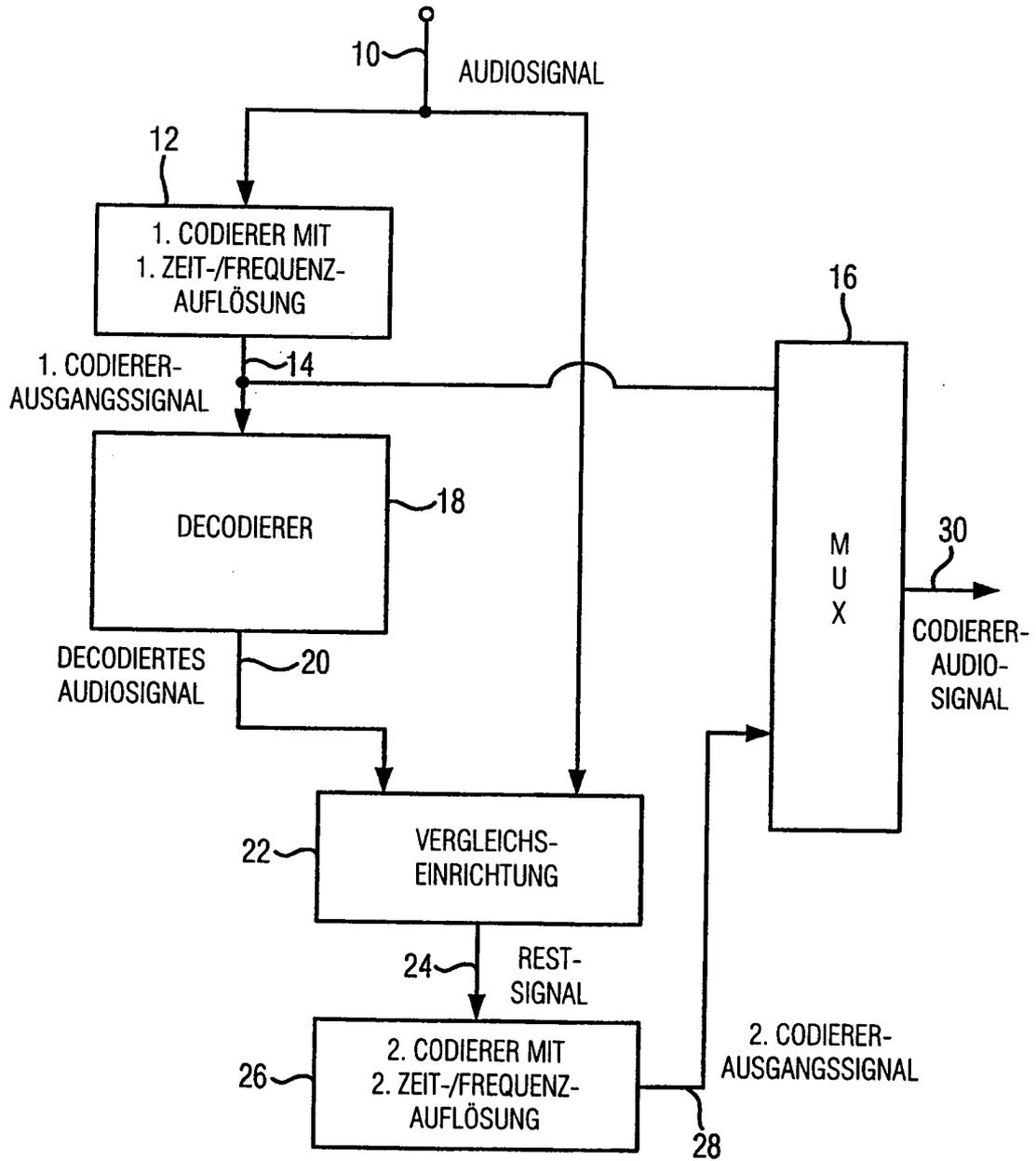
Kombinieren (56) des decodierten Audiosignals und des decodierten Restsignals, um das Ausgangssig-

nal zu erhalten.

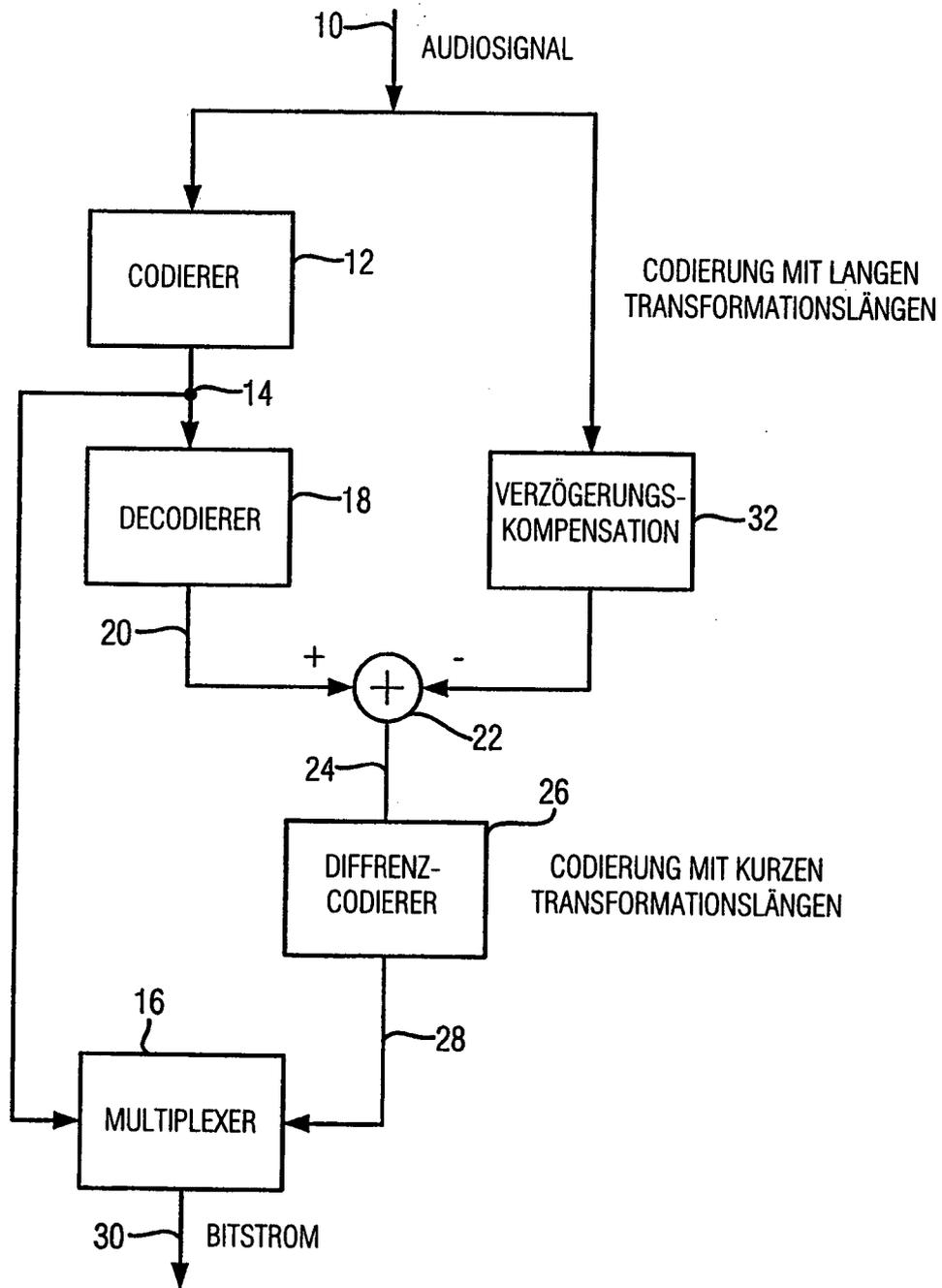
10. Computerprogramm mit einem Programmcode, der alle Schritte des Verfahrens gemäß Anspruch 7 oder 9 durchführt, wenn das Programm auf einem Computer läuft.

Es folgen 4 Blatt Zeichnungen

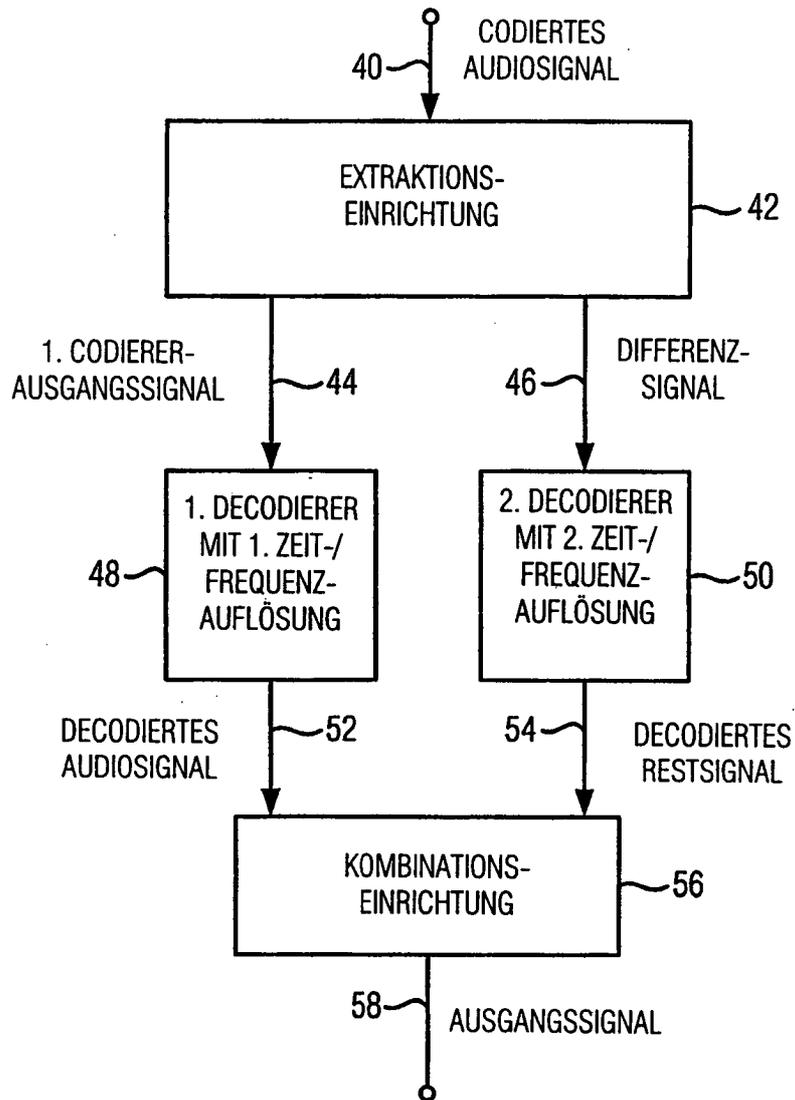
Anhängende Zeichnungen



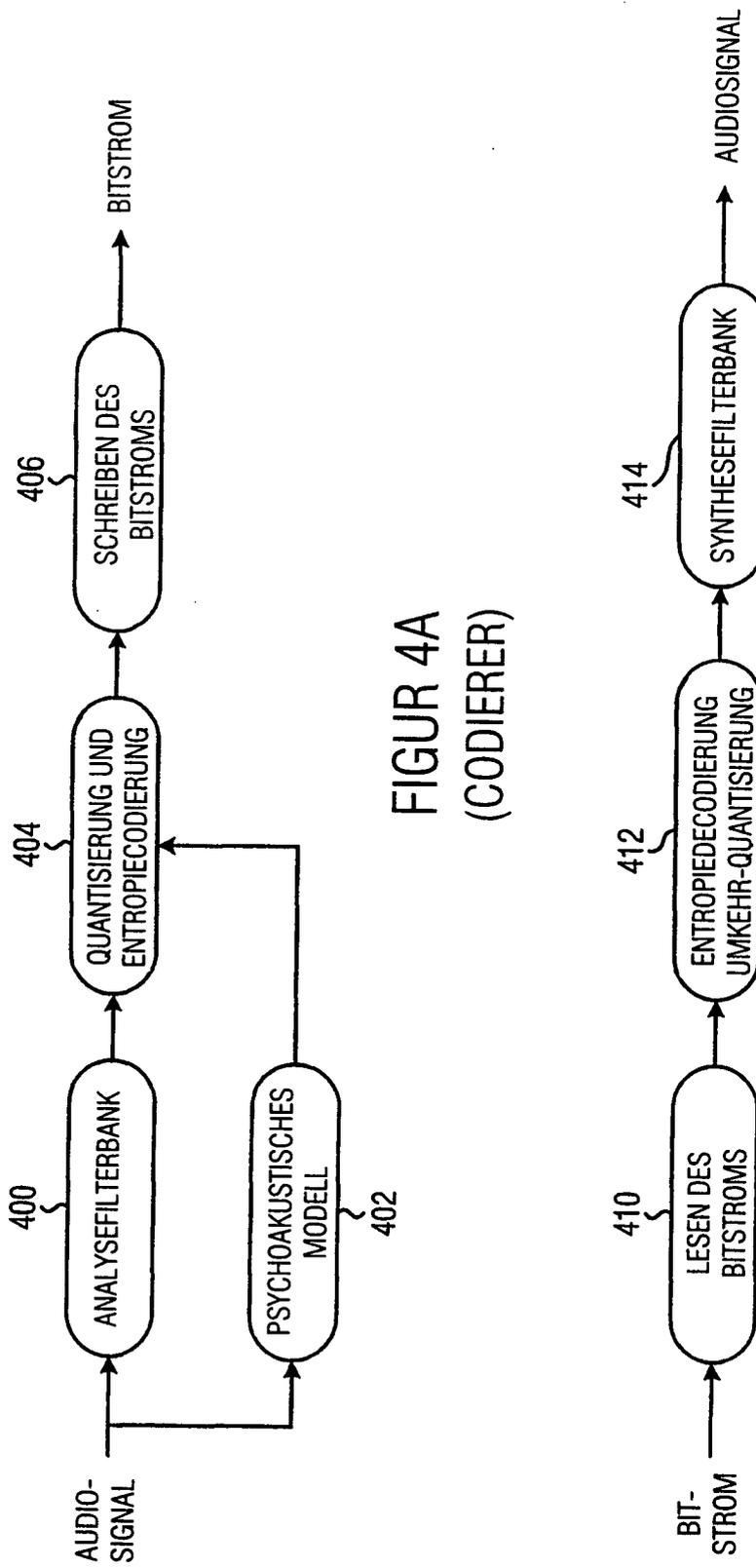
FIGUR 1



FIGUR 2



FIGUR 3



FIGUR 4A
(CODIERER)

FIGUR 4B
(DECODIERER)