



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 10 2004 009 949 A1 2005.09.29**

(12)

Offenlegungsschrift

(21) Aktenzeichen: **10 2004 009 949.9**

(22) Anmeldetag: **01.03.2004**

(43) Offenlegungstag: **29.09.2005**

(51) Int Cl.7: **H04N 7/32**
G10L 19/04

(71) Anmelder:
**Fraunhofer-Gesellschaft zur Förderung der
angewandten Forschung e.V., 80686 München, DE**

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

(72) Erfinder:
**Schug, Michael, 91056 Erlangen, DE; Hilpert,
Johannes, 90411 Nürnberg, DE; Geyersberger,
Stefan, 97076 Würzburg, DE; Neuendorf, Max,
90408 Nürnberg, DE**

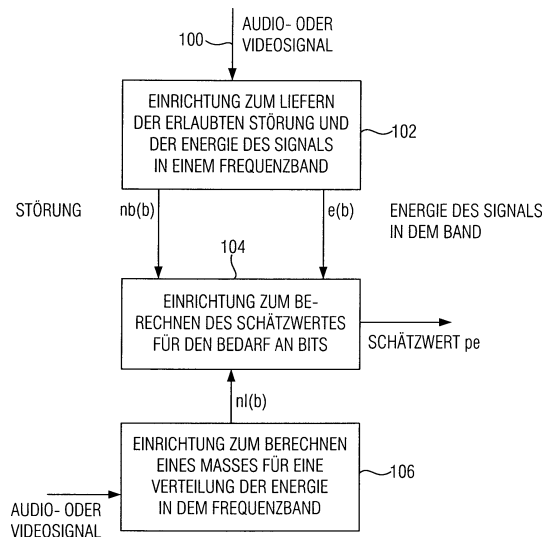
(56) Für die Beurteilung der Patentfähigkeit in Betracht
gezogene Druckschriften:
**James D. Johnston: Estimation of Perceptual
Entropy
Using Noise Masking Criteria;**

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gemäß § 44 PatG ist gestellt.

(54) Bezeichnung: **Vorrichtung und Verfahren zum Ermitteln eines Schätzwertes**

(57) Zusammenfassung: Zum Ermitteln eines Schätzwertes für einen Bedarf an Informationseinheiten zum Codieren eines Signals wird neben der erlaubten Störung für ein Frequenzband und einer Energie des Frequenzbands zusätzlich ein Maß für die Verteilung der Energie in dem Frequenzband berücksichtigt. Damit wird ein besserer Schätzwert für den Bedarf an Informationseinheiten erhalten, so dass effizienter und genauer codiert werden kann.



Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf Codierer zum Codieren eines Signals, das Audio- und/oder Videoinformationen umfasst, und insbesondere auf die Abschätzung für einen Bedarf von Informationseinheiten zum Codieren dieses Signals.

[0002] Nachfolgend wird der bekannte Codierer dargestellt. An einem Eingang **1000** wird ein zu codierendes Audiosignal eingespeist. Dieses wird zunächst einer Skalierungsstufe **1002** zugeführt, in der eine sogenannte AAC-Verstärkungssteuerung durchgeführt wird, um den Pegel des Audiosignals festzulegen. Seiteninformationen aus der Skalierung werden einem Bitstromformatierer **1004** zugeführt, wie es durch den Pfeil zwischen dem Block **1002** und dem Block **1004** dargestellt ist. Das skalierte Audiosignal wird hierauf einer MDCT-Filterbank **1006** zugeführt. Beim AAC-Codierer implementiert die Filterbank eine modifizierte diskrete Cosinustransformation mit 50% überlappenden Fenstern, wobei die Fensterlänge durch einen Block **1008** bestimmt wird.

[0003] Allgemein gesagt ist der Block **1008** dazu vorhanden, dass transiente Signale mit kürzeren Fenstern gefenstert werden, und dass eher stationäre Signale mit längeren Fenstern gefenstert werden. Dies dient dazu, dass aufgrund der kürzeren Fenster für transiente Signale eine höhere Zeitauflösung (auf Kosten der Frequenzauflösung) erreicht wird, während für eher stationäre Signale eine höhere Frequenzauflösung (auf Kosten der Zeitauflösung) durch längere Fenster erreicht wird, wobei tendenziell längere Fenster bevorzugt werden, da sie einen größeren Codiergewinn versprechen. Am Ausgang der Filterbank **1006** liegen zeitlich betrachtet aufeinanderfolgende Blöcke von Spektralwerten vor, die je nach Ausführungsform der Filterbank MDCT-Koeffizienten, Fourier-Koeffizienten oder auch Subbandsignale sein können, wobei jedes Subbandsignal eine bestimmte begrenzte Bandbreite hat, die durch den entsprechenden Subbandkanal in der Filterbank **1006** festgelegt wird, und wobei jedes Subbandsignal eine bestimmte Anzahl von Subband-Abtastwerten aufweist.

[0004] Nachfolgend wird beispielhaft der Fall dargestellt, bei dem die Filterbank zeitlich betrachtet aufeinanderfolgende Blöcke von MDCT-Spektralkoeffizienten ausgibt, die allgemein gesagt, aufeinanderfolgende Kurzzeitspektren des zu codierenden Audiosignals am Eingang **1000** darstellen. Ein Block von MDCT-Spektralwerten wird dann in einen TNS-Verarbeitungsblock **1010** eingespeist, in dem eine zeitliche Rauschformung stattfindet (TNS = temporal noise shaping). Die TNS-Technik wird dazu verwendet, um die zeitliche Form des Quantisierungsrauschens innerhalb jedes Fensters der Transformation zu formen. Dies wird dadurch erreicht, dass ein Filterpro-

zess auf Teile der Spektraldaten jedes Kanals angewendet wird. Die Codierung wird auf einer Fensterbasis durchgeführt. Insbesondere werden die folgenden Schritte ausgeführt, um das TNS-Tool auf ein Fenster spektraler Daten, also auf einen Block von Spektralwerten anzuwenden.

Stand der Technik

[0005] Zunächst wird ein Frequenzbereich für das TNS-Tool ausgewählt. Eine geeignete Auswahl besteht darin, einen Frequenzbereich von 1,5 kHz bis zum höchsten möglichen Skalenfaktorband mit einem Filter abzudecken. Es sei darauf hingewiesen, dass dieser Frequenzbereich von der Abtastrate abhängt, wie es im AAC-Standard (ISO/IEC 14496-3: 2001 (E)) spezifiziert ist.

[0006] Anschließend wird eine LPC-Berechnung (LPC = linear predictive coding = lineare prädiktive Codierung) ausgeführt, und zwar mit den spektralen MDCT-Koeffizienten, die in dem ausgewählten Zielfrequenzbereich liegen. Für eine erhöhte Stabilität werden Koeffizienten, die Frequenzen unter 2,5 kHz entsprechen, aus diesem Prozess ausgeschlossen. Übliche LPC-Prozeduren, wie sie aus der Sprachverarbeitung bekannt sind, können für die LPC-Berechnung verwendet werden, beispielsweise der bekannte Levinson-Durbin-Algorithmus. Die Berechnung wird für die maximal zulässige Ordnung des Rauschformungsfilters ausgeführt.

[0007] Als Ergebnis der LPC-Berechnung wird der erwartete Prädiktionsgewinn PG erhalten. Ferner werden die Reflexionskoeffizienten oder Parcor-Koeffizienten erhalten.

[0008] Wenn der Prädiktionsgewinn eine bestimmte Schwelle nicht überschreitet, wird das TNS-Tool nicht angewendet. In diesem Fall wird eine Steuerinformation in den Bitstrom geschrieben, damit ein Decodierer weiß, dass keine TNS-Verarbeitung ausgeführt worden ist.

[0009] Wenn der Prädiktionsgewinn jedoch eine Schwelle überschreitet, wird die TNS-Verarbeitung angewendet.

[0010] In einem nächsten Schritt werden die Reflexionskoeffizienten quantisiert. Die Ordnung des verwendeten Rauschformungsfilters wird durch Entfernen aller Reflexionskoeffizienten mit einem Absolutwert kleiner als eine Schwelle von dem „Schwanz“ des Reflexionskoeffizienten-Arrays bestimmt.

[0011] Die Anzahl der verbleibenden Reflexionskoeffizienten liegt in der Größenordnung des Rauschformungsfilters. Eine geeignete Schwelle liegt bei 0,1.

[0012] Die verbleibenden Reflexionskoeffizienten werden typischerweise in lineare Prädiktionskoeffizienten umgewandelt, wobei diese Technik auch als "Step-Up"-Prozedur bekannt ist.

[0013] Die berechneten LPC-Koeffizienten werden dann als Codierer-Rauschformungsfilterkoeffizienten, also als Prädiktionsfilterkoeffizienten verwendet. Dieses FIR-Filter wird über den spezifizierten Zielfrequenzbereich geführt. Bei der Decodierung wird ein autoregressives Filter verwendet, während bei der Codierung ein sogenanntes Moving-Average-Filter verwendet wird. Schließlich werden noch die Seiteninformationen für das TNS-Tool dem Bitstromformatierer zugeführt, wie es durch den Pfeil dargestellt ist, der zwischen dem Block TNS-Verarbeitung **1010** und dem Bitstromformatierer **1004** in **Fig. 3** gezeigt ist.

[0014] Hierauf werden mehrere in **Fig. 3** nicht gezeigte optionale Tools durchlaufen, wie beispielsweise ein Langzeitprädiktions-Tool, ein Intensity/Kopplungs-Tool, ein Prädiktions-Tool, ein Rauschsubstitutions-Tool, bis schließlich zu einem Mitte/Seite-Codierer **1012** gelangt wird. Der Mitte/Seite-Codierer **1012** ist dann aktiv, wenn das zu codierende Audiosignal ein Multikanalsignal ist, also ein Stereosignal mit einem linken Kanal und einem rechten Kanal. Bisher, also in der Verarbeitungsrichtung vor dem Block **1012** in **Fig. 3** wurden der linke und der rechte Stereokanal getrennt voneinander verarbeitet, also skaliert, durch die Filterbank transformiert, der TNS-Verarbeitung unterzogen oder nicht etc.

[0015] Im Mitte/Seite-Codierer wird dann zunächst überprüft, ob eine Mitte/Seite-Codierung sinnvoll ist, also überhaupt einen Codiergewinn bringt. Eine Mitte/Seite-Codierung wird dann einen Codiergewinn bringen, wenn der linke und der rechte Kanal eher ähnlich sind, da dann der Mitte-Kanal, also die Summe aus dem linken und dem rechten Kanal nahezu gleich dem linken oder dem rechten Kanal ist, abgesehen von der Skalierung durch den Faktor $1/2$, während der Seite-Kanal nur sehr kleine Werte hat, da er gleich der Differenz zwischen dem linken und dem rechten Kanal ist. Damit ist zu sehen, dass dann, wenn der linke und der rechte Kanal annähernd gleich sind, die Differenz annähernd Null ist bzw. nur ganz kleine Werte umfasst, die – so ist die Hoffnung – in einem nachfolgenden Quantisierer **1014** zu Null quantisiert werden und somit sehr effizient übertragen werden können, da dem Quantisierer **1014** ein Entropie-Codierer **1016** nachgeschaltet ist.

[0016] Dem Quantisierer **1014** wird von einem psycho-akustischen Modell **1020** eine erlaubte Störung pro Skalenfaktorband zugeführt. Der Quantisierer arbeitet iterativ, d. h. es wird zunächst eine äußere Iterationsschleife aufgerufen, die dann eine innere Iterationsschleife aufruft. Allgemein gesagt wird zunächst, ausgehend von Quantisiererschrittwei-

ten-Startwerten, eine Quantisierung eines Blocks von Werten am Eingang des Quantisierers **1014** vorgenommen. Insbesondere quantisiert die innere Schleife die MDCT-Koeffizienten, wobei eine bestimmte Anzahl von Bits verbraucht wird. Die äußere Schleife berechnet die Verzerrung und modifizierte Energie der Koeffizienten unter Verwendung des Skalenfaktors, um wieder eine innere Schleife aufzurufen. Dieser Prozess wird iteriert, bis ein bestimmter Bedingungssatz erfüllt ist. Für jede Iteration in der äußeren Iterationsschleife wird dabei das Signal rekonstruiert, um die durch die Quantisierung eingeführte Störung zu berechnen und mit der von dem psycho-akustischen Modell **1020** gelieferten erlaubten Störung zu vergleichen. Ferner werden die Skalenfaktoren von Iteration zu Iteration um eine Stufe vergrößert, und zwar für jede Iteration der äußeren Iterationsschleife.

[0017] Dann, wenn eine Situation erreicht ist, bei der die durch die Quantisierung eingeführte Quantisierungsstörung unterhalb der durch das psycho-akustische Modell bestimmten erlaubten Störung ist, und wenn gleichzeitig Bitanforderungen erfüllt sind, nämlich, dass eine Maximalbitrate nicht überschritten wird, wird die Iteration, also das Analyse-Durch-Synthese-Verfahren beendet, und es werden die erhaltenen Skalenfaktoren codiert, wie es in dem Block **1014** ausgeführt ist und in codierter Form dem Bitstromformatierer **1004** zugeführt, wie es durch den Pfeil gekennzeichnet ist, der zwischen dem Block **1014** und dem Block **1004** gezeichnet ist. Die quantisierten Werte werden dann dem Entropie-Codierer **1016** zugeführt, der typischerweise unter Verwendung mehrerer Huffman-Code-Tabellen für verschiedene Skalenfaktorbander eine Entropie-Codierung durchführt, um die quantisierten Werte in ein binäres Format zu übertragen. Wie es bekannt ist, wird bei der Entropie-Codierung in Form der Huffman-Codierung auf Code-Tabellen zurückgegriffen, die aufgrund einer erwarteten Signalstatistik erstellt werden, und bei denen häufig auftretende Werte kürzere Code-Wörter bekommen als seltener auftretende Werte. Die entropiecodierten Werte werden dann ebenfalls als eigentliche Hauptinformationen dem Bitstromformatierer **1004** zugeführt, der dann gemäß einer bestimmten Bitstromsyntax ausgangsseitig das codierte Audiosignal ausgibt.

[0018] Die Datenreduktion von Audiosignalen ist mittlerweile eine bekannte Technik, die Gegenstand einer Reihe von Internationalen Standards ist (z. B. ISO/MPEG-1, MPEG-2 AAC, MPEG-4).

[0019] Gemeinsam ist den oben genannten Verfahren, dass das Eingangssignal mittels eines sogenannten Encoders unter Ausnutzung wahrnehmungsbezogener Effekte (Psychoakustik, Psychooptik) in eine kompakte, datenreduzierte Darstellung gebracht wird. Hierzu wird üblicherweise eine Spektralanalyse des Signals vorgenommen und die ent-

sprechende Signalkomponenten werden unter Berücksichtigung eines Wahrnehmungsmodells quantisiert und anschließend in möglichst kompakter Weise als sogenannter Bitstrom codiert.

[0020] Um vor der eigentlichen Quantisierung abzuschätzen, wie viele Bits ein bestimmter zu codierender Abschnitt des Signals benötigen wird, kann die sogenannte Perceptual Entropy (PE) herangezogen werden. Die PE liefert auch ein Maß dafür, wie schwierig es für den Encoder ist, ein bestimmtes Signal oder Teile davon zu codieren.

[0021] Entscheidend für die Qualität der Abschätzung ist die Abweichung der PE von der Anzahl tatsächlich benötigter Bits.

[0022] Ferner kann die Perceptual Entropy bzw. jeder Schätzwert für einen Bedarf von Informationseinheiten zum Codieren eines Signals dafür herangezogen werden, abzuschätzen, ob das Signal transient oder stationär ist, da transiente Signale ebenfalls mehr Bits zum Codieren benötigen als eher stationäre Signale. Die Abschätzung einer transienten Eigenschaft eines Signal wird beispielsweise dazu verwendet, um eine Fensterlängenentscheidung, wie sie um Block **1008** in **Fig. 3** angedeutet ist, durchzuführen.

[0023] In **Fig. 6** ist die Perceptual Entropy berechnet nach ISO/IEC IS 13818-7 (MPEG-2 advanced audio coding (AAC)) dargestellt. Zu Berechnung dieser Perceptual Entropy, also einer bandweisen Perceptual Entropy wird die in **Fig. 6** dargestellte Gleichung verwendet. In dieser Gleichung steht der Parameter p_e für die Perceptual Entropy. Ferner steht $\text{width}(b)$ für die Anzahl der Spektralkoeffizienten im jeweiligen Band b . Ferner ist $e(b)$ die Energie des Signals in diesem Band. Schließlich ist $n_b(b)$ die dazu passende Maskierungsschwelle bzw. allgemeiner ausgedrückt, die erlaubte Störung, die in das Signal eingebracht werden kann, beispielsweise durch eine Quantisierung, damit ein menschlicher Hörer dennoch keine oder nur eine verschwindend geringe Störung hört.

[0024] Die Bänder können von der Bändeinteilung des psychoakustischen Modells (Block **1020** in **Fig. 3**) stammen, oder es handelt sich um die bei der Quantisierung verwendeten sogenannten Skalenfaktorbänder ($scfb$). Die psychoakustische Maskierungsschwelle ist der Energiewert, den der Quantisierungsfehler nicht überschreiten sollte.

[0025] Die in **Fig. 6** gezeigte Abbildung zeigt somit, wie gut eine so bestimmte Perceptual Entropy ab Abschätzung für die Anzahl der zur Codierung benötigten Bits funktioniert. Hierzu wurde am Beispiel eines AAC-Codierers bei unterschiedlichen Bitraten für jeden einzelnen Block die jeweilige Perceptual Entropy in Abhängigkeit von den verbrauchten Bits aufgetragen. Das verwendete Teststück beinhaltet eine typi-

sche Mischung aus Musik, Sprache und Einzelinstrumenten.

[0026] Idealerweise würden sich die Punkte entlang einer Geraden durch den Nullpunkt versammeln. Die Ausdehnung der Punktfolge mit den Abweichungen von der idealen Linie verdeutlicht die ungenaue Abschätzung.

[0027] Nachteilig an dem in **Fig. 6** gezeigten Konzept ist also die Abweichung, die sich dahin gehend äußert, dass sich z. B. ein zu großer Wert für die Perceptual Entropy ergibt, was wiederum bedeutet, dass dem Quantisierer signalisiert wird, dass mehr Bits als eigentlich erforderlich, benötigt werden. Dies führt dazu, dass der Quantisierer zu fein quantisiert, dass er also nicht das Maß an erlaubter Störung ausschöpft, was in einem reduzierten Codiergewinn resultiert. Andererseits, wenn der Wert für die Perceptual Entropy zu klein ermittelt wird, so wird dem Quantisierer signalisiert, dass weniger Bits als eigentlich erforderlich, zur Codierung des Signals benötigt werden. Dies wiederum hat zur Folge, dass der Quantisierer zu grob quantisiert, was unmittelbar zu einer hörbaren Störung im Signal führen würde, sofern nicht Gegenmaßnahmen ergriffen werden. Die Gegenmaßnahmen können darin bestehen, dass der Quantisierer noch eine oder mehrere weitere Iterationsschleifen benötigt, was die Rechenzeit des Codierers ansteigen lässt.

[0028] Zur Verbesserung der Berechnung der Perceptual Entropy könnte man, wie es in **Fig. 7** gezeigt ist, einen konstanten Term, wie beispielsweise 1,5, in den Logarithmus-Ausdruck einführen. Dann ergibt sich bereits ein besseres Ergebnis, also eine geringere Abweichung nach oben bzw. unten, obgleich dennoch zu sehen ist, dass bei der Berücksichtigung eines konstanten Terms im Logarithmus-Ausdruck zwar der Fall reduziert ist, dass die Perceptual Entropy einen zu optimistischen Bedarf an Bits signalisiert. Andererseits ist aus **Fig. 7** jedoch deutlich zu erkennen, dass signifikant eine zu hohe Anzahl an Bits signalisiert wird, was dazu führt, dass der Quantisierer immer zu fein quantisieren wird, dass also der Bitbedarf größer angenommen wird, als er eigentlich ist, was wiederum in einem reduzierten Codiergewinn resultiert. Die Konstante in dem Logarithmus-Ausdruck ist eine grobe Abschätzung der für die Seiteninformationen benötigten Bits.

[0029] So liefert das Einfügen eines Terms in den Logarithmus-Ausdruck zwar eine Verbesserung der bandweisen Perceptual Entropy, wie es in **Fig. 6** dargestellt ist, da die Bänder mit sehr geringem Abstand zwischen Energie und Maskierungsschwelle eher berücksichtigt werden, da auch für die Übertragung von zu Null quantisierten Spektralkoeffizienten eine gewisse Anzahl von Bits nötig ist.

[0030] Eine weitere, jedoch sehr Rechenzeit-aufwendige Berechnung der Perceptual Entropy ist in **Fig. 8** dargestellt. In **Fig. 8** ist der Fall gezeigt, bei dem die Perceptual Entropy linienweise berechnet wird. Der Nachteil liegt jedoch in dem höheren Rechenaufwand der linienweisen Berechnung. Hier werden anstelle der Energie Spektralkoeffizienten $X(k)$ eingesetzt, wobei $k_{\text{Offset}}(b)$ den ersten Index von Band b bezeichnet. Wenn **Fig. 8** mit **Fig. 7** verglichen wird, so ist deutlich im Bereich zwischen 2000 und 3000 Bit eine Reduzierung der „Ausschläge“ nach oben zu erkennen. Die PE-Schätzung wird daher genauer sein, also nicht zu pessimistisch schätzen, sondern eher am Optimum liegen, so dass der Codiergewinn im Vergleich zu den in **Fig. 6** und **7** gezeigten Berechnungsverfahren ansteigen kann, bzw. die Anzahl der Iterationen im Quantisierer wird reduziert.

[0031] Nachteilig an der linienweise Berechnung der Perceptual Entropy ist jedoch die Rechenzeit, die benötigt wird, um die in **Fig. 8** gezeigte Gleichung auszuwerten.

[0032] So spielen solche Rechenzeitennachteile zwar nicht unbedingt eine Rolle, wenn der Codierer auf einem leistungsstarken PC oder einer leistungsstarken Workstation läuft. Ganz anders sieht es dagegen aus, wenn der Codierer in einem tragbaren Gerät, wie beispielsweise einem UMTS-Handy untergebracht ist, das einerseits klein und billig sein muss, das andererseits einen niedrigen Strombedarf haben muss, und das zusätzlich schnell arbeiten muss, um die Codierung eines über die UMTS-Verbindung übertragenen Audiosignals oder Videosignals zu ermöglichen.

Aufgabenstellung

[0033] Die Aufgabe der vorliegenden Erfindung besteht darin, ein effizientes und dennoch genaues Konzept zum Ermitteln eines Schätzwerts für einen Bedarf von Informationseinheiten zum Codieren eines Signals zu schaffen.

[0034] Diese Aufgabe wird durch eine Vorrichtung gemäß Patentanspruch 1, ein Verfahren gemäß Patentanspruch 12 oder ein Computerprogramm nach Patentanspruch 13 gelöst.

[0035] Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, dass an einer frequenzbandweisen Berechnung des Schätzwerts für einen Bedarf an Informationseinheiten aus Rechenzeitgründen festgehalten werden muss, dass jedoch, um eine genaue Ermittlung des Schätzwerts zu erhalten, die Verteilung der Energie in dem Frequenzband, das bandweise zu berechnen ist, berücksichtigt werden muss.

[0036] Damit wird gewissermaßen implizit der dem

Quantisierer nachfolgende Entropie-Codierer in die Ermittlung des Schätzwerts für den Bedarf von Informationseinheiten „hineingezogen“. Die Entropy-Codierung ermöglicht es nämlich, dass zur Übertragung von kleineren Spektralwerten eine geringere Anzahl an Bits benötigt wird als zur Übertragung von größeren Spektralwerten. Besonders effizient ist der Entropie-Codierer dann, wenn zu-Null-quantisierte Spektralwerte übertragen werden können. Da diese typischerweise am häufigsten auftreten werden, ist das Codewort zum Übertragen einer zu-Null-quantisierten Spektrallinie das kürzeste Codewort, und ist das Codewort zum Übertragen einer immer größeren quantisierten Spektrallinie immer länger. Darüber hinaus kann für ein besonders effizientes Konzept zum Übertragen einer Folge von zu-Null-quantisierten Spektralwerten sogar auf eine Lauflängencodierung zurückgegriffen werden, was zur Folge hat, dass im Falle eines Laufs von Nullen pro zu-Null-quantisiertem Spektralwert durchschnittlich betrachtet nicht einmal ein einziges Bit benötigt wird.

[0037] Es wurde herausgefunden, dass die im Stand der Technik verwendete bandweise Perceptual-Entropy-Berechnung zur Ermittlung des Schätzwerts für den Bedarf von Informationseinheiten die Wirkungsweise des nachgeschalteten Entropie-Codierers völlig ignoriert, wenn die Verteilung der Energie in dem Frequenzband von einer vollständig gleichmäßigen Verteilung abweicht.

[0038] Erfindungsgemäß wird somit zur Reduktion der Ungenauigkeiten der bandweisen Berechnung berücksichtigt, wie die Energie innerhalb eines Bandes verteilt ist.

[0039] Je nach Implementierung kann das Maß für die Verteilung der Energie in dem Frequenzband auf der Basis der tatsächlichen Amplituden ermittelt werden, oder durch eine Schätzung der Frequenzlinien, die durch den Quantisierer nicht zu null quantisiert werden. Dieses Maß, das auch als „nl“ bezeichnet wird, wobei nl für „number of active lines“, also für die Anzahl von aktiven Linien, steht, wird aus Rechenzeit-Effizienzgründen bevorzugt. Es kann jedoch auch die Anzahl der zu null quantisierten Spektrallinien oder eine feinere Unterteilung berücksichtigt werden, wobei diese Schätzung immer genauer wird, je mehr Informationen des nachgeschalteten Entropie-Codierers berücksichtigt werden. Ist der Entropie-Codierer auf der Basis von Huffman-Codetabellen aufgebaut, so können Eigenschaften dieser Codetabellen besonders gut integriert werden, da die Codetabellen nicht aufgrund der Signalstatistik gewissermaßen on-line berechnet werden, sondern da die Codetabellen unabhängig von dem tatsächlichen Signal ohnehin feststehen.

[0040] Je nach Rechenzeit-Einschränkungen wird jedoch im Falle einer besonders effizienten Berech-

nung das Maß für die Verteilung der Energie in dem Frequenzband durch die Ermittlung der nach der Quantisierung noch überlebenden Linien, also der Anzahl von aktiven Linien, durchgeführt.

[0041] Die vorliegende Erfindung ist dahingehend vorteilhaft, dass ein Schätzwert für einen Bedarf an Informationsinhalten ermittelt wird, der zum einen genauer und zum anderen effizienter als im Stand der Technik ist.

[0042] Darüber hinaus ist die vorliegende Erfindung für verschiedene Anwendungen skalierbar, da je nach erwünschter Genauigkeit des Schätzwerts immer mehr Eigenschaften des Entropie-Codierers, jedoch zum Preis einer erhöhten Rechenzeit, in die Schätzung des Bitbedarfs mit hereingenommen werden können.

Ausführungsbeispiel

[0043] Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend bezugnehmend auf die beiliegenden Zeiten detailliert erläutert. Es zeigen:

[0044] [Fig. 1](#) ein Blockschaltbild der erfindungsgemäßen Vorrichtung zum Ermitteln eines Schätzwerts;

[0045] [Fig. 2a](#) eine bevorzugte Ausführungsform der Einrichtung zum Berechnen eines Maßes für die Verteilung der Energie in dem Frequenzband;

[0046] [Fig. 2b](#) eine bevorzugte Ausführungsform der Einrichtung zum Berechnen des Schätzwerts für den Bedarf an Bits;

[0047] [Fig. 3](#) ein Blockschaltbild eines bekannten Audio-Codierers;

[0048] [Fig. 4](#) eine Prinzipdarstellung zur Erläuterung des Einflusses der Energieverteilung innerhalb eines Bandes auf die Ermittlung des Schätzwerts;

[0049] [Fig. 5](#) ein Diagramm zur Schätzwertberechnung gemäß der vorliegenden Erfindung;

[0050] [Fig. 6](#) ein Diagramm zur Schätzwertberechnung gemäß ISO/IEC IS 13818-7(AAC);

[0051] [Fig. 7](#) ein Diagramm zur Schätzwertberechnung mit konstantem Term;

[0052] [Fig. 8](#) ein Diagramm zur linienweisen Schätzwertberechnung mit konstantem Term.

[0053] Nachfolgend wird bezugnehmend auf [Fig. 1](#) die erfindungsgemäße Vorrichtung zum Ermitteln eines Schätzwerts für einen Bedarf von Informationseinheiten zum Codieren eines Signals dargestellt.

Das Signal, das ein Audio- und/oder ein Videosignal sein kann, wird über einen Eingang **100** eingespeist. Vorzugsweise liegt das Signal bereits als spektrale Darstellung mit Spektralwerten vor. Dies ist jedoch nicht unbedingt erforderlich, da durch entsprechende z. B. Bandpass-Filterung auch einige Berechnungen mit einem Zeitsignal durchgeführt werden können.

[0054] Das Signal wird einer Einrichtung **102** zum Liefern eines Maßes für eine erlaubte Störung für ein Frequenzband des Signals zugeführt. Die erlaubte Störung kann beispielsweise mittels eines psycho-akustischen Modells, wie es anhand von [Fig. 3](#) (Block **1020**) erläutert worden ist, ermittelt werden. Die Einrichtung **102** ist ferner wirksam, um auch ein Maß für die Energie des Signals in dem Frequenzband zu liefern. Voraussetzung für eine bandweise Berechnung ist, dass ein Frequenzband, für das eine erlaubte Störung oder eine Signalenergie angegeben wird, wenigstens zwei oder mehrere Spektrallinien der spektralen Darstellung des Signals enthält. Bei typischen standardisierten Audio-Codierern wird das Frequenzband vorzugsweise ein Skalenfaktorband sein, da die Bitbedarfsschätzung unmittelbar vom Quantisierer benötigt wird, um festzustellen, ob eine erfolgte Quantisierung ein Bitkriterium erfüllt oder nicht.

[0055] Die Einrichtung **102** ist ausgebildet, um sowohl die erlaubte Störung $nb(b)$, als auch die Signalenergie $e(b)$ des Signals in dem Band einer Einrichtung **104** zum Berechnen des Schätzwerts für den Bedarf an Bits zuzuführen.

[0056] Erfindungsgemäß ist die Einrichtung **104** zum Berechnen des Schätzwerts für den Bedarf von Bits ausgebildet, um neben der erlaubten Störung und der Signalenergie ein Maß $nl(b)$ für eine Verteilung der Energie in dem Frequenzband zu berücksichtigen, wobei die Verteilung der Energie in dem Frequenzband von einer vollständig gleichmäßigen Verteilung abweicht. Das Maß für die Verteilung der Energie wird in einer Einrichtung **106** berechnet, wobei die Einrichtung **106** zumindest ein Band, nämlich das betrachtete Frequenzband des Audio- oder Videosignals entweder als Bandpass-Signal oder direkt als Folge von Spektrallinien benötigt, um z. B. eine spektrale Analyse des Bandes durchführen zu können, um das Maß für die Verteilung der Energien im Frequenzband zu erhalten.

[0057] Selbstverständlich kann das Audio- oder Videosignal der Einrichtung **106** als Zeitsignal zugeführt werden, wobei die Einrichtung **106** dann eine Bandfilterung sowie eine Analyse in dem Band durchführt. Alternativ kann das Audio- oder Videosignal, das der Einrichtung **106** zugeführt wird, bereits im Frequenzbereich vorliegen, wie z. B. als MDCT-Koeffizienten, oder aber auch als Bandpass-Signal in der Filterbank mit einer im Vergleich zu einer MDCT-Fil-

terbank kleineren Anzahl an Bandpass-Filtern.

[0058] Bei einem bevorzugten Ausführungsbeispiel ist die Einrichtung **106** zum Berechnen ausgebildet, um zur Berechnung des Schätzwerts aktuelle Beträge von Spektralwerten in dem Frequenzband zu berücksichtigen.

[0059] Ferner kann die Einrichtung zum Berechnen des Maßes für die Verteilung der Energie ausgebildet sein, um als Maß für die Verteilung der Energie eine Anzahl von Spektralwerten zu ermitteln, deren Betrag größer oder gleich einer vorbestimmten Betragsschwelle sind, oder deren Betrag kleiner oder gleich der Betragsschwelle ist, wobei die Betragsschwelle vorzugsweise eine geschätzte Quantisiererstufe ist, die in einem Quantisierer bewirkt, dass Werte kleiner oder gleich der Quantisiererstufe zu null quantisiert werden. In diesem Fall ist das Maß für die Energie die Anzahl von aktiven Linien, also die Anzahl der Linien, die nach der Quantisierung überleben.

[0060] [Fig. 2a](#) zeigt ein bevorzugtes Ausführungsbeispiel für die Einrichtung **106** zum Berechnen des Maßes für die Verteilung der Energie in dem Frequenzband. Das Maß für die Verteilung der Energie in dem Frequenzband ist in [Fig. 2a](#) mit $l(b)$ bezeichnet. Der Formfaktor $ffac(b)$ ist bereits ein Maß für die Verteilung der Energie in dem Frequenzband. Wie es aus Block **106** ersichtlich ist, wird das Maß für die spektrale Verteilung nl aus dem Formfaktor $ffac(b)$ durch Gewichtung mit der 4. Wurzel aus der Signalenergie $e(b)$ geteilt durch die Bandbreite b ermittelt.

[0061] Der Formfaktor $ffac(b)$ errechnet sich durch Betragsbildung einer Spektrallinie und anschließender Wurzelbildung dieser Spektrallinie und anschließender Aufsummierung der „gewurzelten“ Beträge der Spektrallinien in dem Band.

[0062] [Fig. 2b](#) zeigt eine bevorzugte Ausführungsform der Einrichtung **104** zum Berechnen des Schätzwerts pe , wobei in [Fig. 2b](#) noch eine Fallunterscheidung eingeführt ist, nämlich dann, wenn der Logarithmus zur Basis 2 des Verhältnisses aus der Energie zur erlaubten Störung größer als ein konstanter Faktor $c1$ oder gleich dem konstanten Faktor ist. In diesem Fall wird die in dem Block **104** oben stehende Alternative genommen, also das Maß für die spektrale Verteilung nl wird mit dem Logarithmusausdruck multipliziert.

[0063] Wird dagegen festgestellt, dass der Logarithmus zur Basis 2 aus dem Verhältnis der Signalenergie zur erlaubten Störung kleiner als der Wert $c1$ ist, so wird die untere Alternative im Block **104** von [Fig. 2b](#) verwendet, die zusätzlich noch eine additive Konstante $c2$ sowie eine multiplikative Konstante $c3$ aufweist, die sich aus den Konstanten $c2$ und $c1$ berechnet.

[0064] Nachfolgend wird anhand von [Fig. 4a](#) und [Fig. 4b](#) das erfindungsgemäße Konzept dargestellt. So zeigt [Fig. 4a](#) ein Band, in dem vier Spektrallinien vorhanden sind, die alle gleich groß sind. Die Energie in diesem Band ist somit gleichmäßig über das Band verteilt. Dagegen zeigt [Fig. 4b](#) eine Situation, bei der die Energie in dem Band in einer Spektrallinie residiert, während die anderen drei Spektrallinien gleich null sind. Das in [Fig. 4b](#) gezeigte Band könnte beispielsweise vor der Quantisierung vorliegen, oder könnte nach der Quantisierung erhalten werden, wenn die in [Fig. 4b](#) zu null gesetzten Spektrallinien vor der Quantisierung kleiner als die erste Quantisiererstufe sind und somit durch den Quantisierer zu null gesetzt werden, also nicht „überleben“.

[0065] Die Anzahl von aktiven Linien in [Fig. 4b](#) ist somit gleich 1, wobei der Parameter nl in [Fig. 4b](#) zu der Quadratwurzel von 2 berechnet wird. Dagegen wird der Wert nl , also das Maß für die spektrale Verteilung der Energie in [Fig. 4a](#) zu 4 berechnet. Dies bedeutet, dass die spektrale Verteilung der Energie gleichmäßiger ist, wenn das Maß für die Verteilung der spektralen Energie größer ist.

[0066] Es sei darauf hingewiesen, dass die bandweise Berechnung der Perceptual Entropy gemäß dem Stand der Technik keinen Unterschied zwischen den beiden Fällen feststellt. Insbesondere wird kein Unterschied festgestellt, wenn in den beiden Bändern, die in [Fig. 4a](#) und [Fig. 4b](#) gezeigt sind, dieselbe Energie vorhanden ist.

[0067] Offensichtlich ist jedoch der in [Fig. 4b](#) gezeigte Fall mit nur einer relevanten Linie mit weniger Bits codierbar, da die drei zu null gesetzten Spektrallinien sehr effizient übertragen werden können. Allgemein gesagt beruht die einfachere Quantisierbarkeit des in [Fig. 4b](#) gezeigten Falls auf der Tatsache, dass nach der Quantisierung und verlustlosen Codierung kleinere Werte und insbesondere zu null quantisierte Werte weniger Bits zur Übertragung benötigen.

[0068] Erfindungsgemäß wird somit berücksichtigt, wie die Energie innerhalb des Bands verteilt ist. Dies erfolgt, wie es ausgeführt worden ist, durch Ersetzen der Anzahl der Linien pro Band in der bekannten Gleichung ([Fig. 6](#)) durch eine Abschätzung der Anzahl der Linien, die nach der Quantisierung ungleich null sind. Diese Abschätzung ist in [Fig. 2a](#) gezeigt.

[0069] Ferner sei darauf hingewiesen, dass der in [Fig. 2a](#) gezeigte Formfaktor auch an anderer Stelle im Codierer benötigt wird, beispielsweise innerhalb des Quantisierungsblocks **1014** zur Bestimmung der Quantisierungs-Schrittweite. Dann, wenn der Formfaktor bereits an anderer Stelle berechnet wird, muß er zur Bit-Abschätzung nicht erneut berechnet werden, so dass das erfindungsgemäße Konzept zur verbesserten Abschätzung des Maßes für die benö-

tigten Bits mit einem Minimum an zusätzlichem Rechenaufwand auskommt.

[0070] Wie es bereits ausgeführt worden ist, handelt es sich bei $X(k)$ um den später zu quantisierenden Spektralkoeffizienten, während die Variable $k_{\text{Offset}}(b)$ den ersten Index im Band b bezeichnet.

[0071] Wie es aus [Fig. 4a](#) und [Fig. 4b](#) ersichtlich ist, ergibt das Spektrum in [Fig. 4a](#) einen Wert $n_1 = 4$, während das Spektrum in [Fig. 4b](#) einen Wert von 1,41 ergibt. Mit Hilfe des Formfaktors steht somit ein Maß für die Charakterisierung der spektralen Feldstruktur innerhalb des Bandes zur Verfügung.

[0072] Die neue Formel zur Berechnung einer verbesserten bandweisen Perceptual Entropie basiert somit auf der Multiplikation des Maßes für die spektrale Verteilung der Energie und des Logarithmus-Ausdrucks, indem die Signalenergie $e(b)$ im Zähler und die erlaubte Störung im Nenner auftreten, wobei je nach Bedarf ein Term innerhalb des Logarithmus eingesetzt werden kann, wie es bereits in [Fig. 7](#) dargestellt ist. Diese Term kann beispielsweise ebenfalls 1,5 sein, kann jedoch auch gleich null sein, wie in dem in [Fig. 2b](#) gezeigten Fall, wobei dies z. B. empirisch bestimmt werden kann.

[0073] An dieser Stelle sei nochmals auf [Fig. 5](#) hingewiesen, aus der die erfindungsgemäß berechnete Perceptual Entropie ersichtlich ist, und zwar aufgetragen über den benötigten Bits. Eine höhere Genauigkeit der Abschätzung gegenüber den Vergleichsbeispielen in den [Fig. 6](#), [Fig. 7](#) und [Fig. 8](#) ist deutlich zu erkennen. Auch gegenüber der linienweisen Berechnung schneidet die erfindungsgemäße modifizierte bandweise Berechnung zumindest gleichwertig ab.

[0074] Abhängig von der Gegebenheit, kann das erfindungsgemäße Verfahren in Hardware oder in Software implementiert werden. Die Implementierung kann auf einem digitalen Speichermedium, insbesondere einer Diskette oder CD mit elektronisch auslesbaren Steuersignalen erfolgen, die so mit einem programmierbaren Computersystem zusammenwirken können, dass das Verfahren ausgeführt wird. Allgemein besteht die Erfindung somit auch in einem Computer-Programm-Produkt mit einem auf einem maschinenlesbaren Träger gespeicherten Programmcode zur Durchführung des erfindungsgemäßen Verfahrens, wenn das Computer-Programm-Produkt auf einem Rechner abläuft. In anderen Worten ausgedrückt, kann die Erfindung somit als ein Computer-Programm mit einem Programmcode zur Durchführung des Verfahrens realisiert werden, wenn das Computer-Programm auf einem Computer abläuft.

Patentansprüche

1. Vorrichtung zum Ermitteln eines Schätzwerts für einen Bedarf an Informationseinheiten zum Codieren eines Signals, das Audio- oder Videoinformationen aufweist mit folgenden Merkmalen: einer Einrichtung (**102**) zum Liefern eines Maßes für eine erlaubte Störung für ein Frequenzband des Signals, wobei das Frequenzband wenigstens zwei Spektralwerte einer spektralen Darstellung des Signals umfasst, und eines Maßes für eine Energie des Signals in dem Frequenzband; einer Einrichtung (**106**) zum Berechnen eines Maßes für eine Verteilung der Energie in dem Frequenzband, wobei die Verteilung der Energie in dem Frequenzband von einer vollständig gleichmäßigen Verteilung abweicht; und einer Einrichtung (**104**) zum Berechnen des Schätzwerts unter Verwendung des Maßes für die Störung, des Maßes für die Energie und des Maßes für die Verteilung der Energie.

2. Vorrichtung nach Anspruch 1, bei der die Einrichtung (**106**) zum Berechnen ausgebildet ist, um zur Berechnung des Maßes für die Verteilung der Energie Beträge von Spektralwerten in dem Frequenzband zu berücksichtigen.

3. Vorrichtung nach Anspruch 1 oder 2, bei der die Einrichtung (**106**) zum Berechnen des Maßes für die Verteilung der Energie ausgebildet ist, um als Maß für die Verteilung der Energie eine Anzahl von Spektralwerten zu ermitteln, deren Betrag größer oder gleich einer vorbestimmten Betragsschwelle sind, oder deren Betrag kleiner oder gleich der Betragsschwelle sind.

4. Vorrichtung nach Anspruch 3, bei der die Betragsschwelle eine exakte oder geschätzte Quantisierstufe ist, die in einem Quantisierer bewirkt, dass Werte kleiner oder gleich der Quantisierstufe zu null quantisiert werden.

5. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (**106**) zum Berechnen ausgebildet ist, um einen Formfaktor gemäß folgender Gleichung zu berechnen:

$$ffac(b) = \sum_{k=k_{\text{Offset}}(b)}^{k_{\text{Offset}}(b+1)-1} \sqrt{|X(k)|},$$

wobei $X(k)$ ein Spektralwert bei einem Frequenzindex k ist, wobei k_{Offset} ein erster Spektralwert in einem Band b ist, und wobei $ffac(b)$ der Formfaktor ist.

6. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (**106**) zum Berechnen ausgebildet ist, um eine vierte Wurzel aus einem Verhältnis zwischen der Energie in dem Frequenzband und einer Breite des Frequenzbands zu

berücksichtigen.

7. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (106) zum Berechnen ausgebildet ist, um das Maß für die Verteilung der Energie gemäß folgender Gleichungen zu berechnen:

$$nl(b) = \frac{ffac(b)}{\left(\frac{e(b)}{width(b)}\right)^{0.25}}$$

$$ffac(b) = \sum_{k=kOffset(b)}^{kOffset(b+1)-1} \sqrt{|X(k)|},$$

wobei $X(k)$ ein Spektralwert bei einem Frequenzindex k ist, wobei $kOffset$ ein erster Spektralwert in einem Band b ist, wobei $ffac(b)$ ein Formfaktor ist, wobei $nl(b)$ das Maß für die Verteilung der Energie in dem Band b darstellt, wobei $e(b)$ eine Signalenergie in dem Band b ist, und wobei $width(b)$ eine Breite des Bandes ist.

8. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (104) zum Berechnen des Schätzwerts ausgebildet ist, um einen Quotienten aus der Energie in dem Frequenzband und der Störung in dem Frequenzband zu verwenden.

9. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (104) zum Berechnen des Schätzwerts ausgebildet ist, um den Schätzwert unter Verwendung des folgenden Ausdrucks zu berechnen:

$$pe = \sum_b nl(b) \cdot \log_2 \left(\frac{e(b)}{nb(b)} + 1.5 \right)$$

wobei pe der Schätzwert ist, wobei $nl(b)$ das Maß für die Verteilung der Energie in dem Band b darstellt, wobei $e(b)$ eine Energie des Signals in dem Band b ist, wobei $nb(b)$ die erlaubte Störung in dem Band b ist, und wobei s ein additiver Term ist, der vorzugsweise gleich 1,5 ist.

10. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (104) zum Berechnen des Schätzwerts ausgebildet ist, um den Schätzwert gemäß folgender Gleichung zu berechnen:

$$pe = \sum_b nl(b) \cdot \log_2 \left(\frac{e(b)}{nb(b)} + 1.5 \right),$$

wobei gilt:

$$nl(b) = \frac{ffac(b)}{\left(\frac{e(b)}{width(b)}\right)^{0.25}}, \text{ und}$$

wobei gilt:

$$ffac(b) = \sum_{k=kOffset(b)}^{kOffset(b+1)-1} \sqrt{|X(k)|},$$

wobei pe der Schätzwert ist, wobei $nl(b)$ das Maß für die Verteilung der Energie in dem Band b darstellt, wobei $e(b)$ eine Energie des Signals in dem Band b ist, wobei $nb(b)$ die erlaubte Störung in dem Band b ist, wobei s ein additiver Term ist, der vorzugsweise gleich 1,5 ist, wobei $X(k)$ ein Spektralwert bei einem Frequenzindex k ist, wobei $kOffset$ ein erster Spektralwert in einem Band b ist, wobei $ffac(b)$ ein Formfaktor ist, und wobei $width(b)$ eine Breite des Bandes ist

11. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der das Signal als spektrale Darstellung mit Spektralwerten gegeben ist.

12. Verfahren zum Ermitteln eines Schätzwerts für einen Bedarf an Informationseinheiten zum Codieren eines Signals, das Audio- oder Videoinformationen aufweist mit folgenden Schritten:

Liefere (102) eines Maßes für eine erlaubte Störung für ein Frequenzband des Signals, wobei das Frequenzband wenigstens zwei Spektralwerte einer spektralen Darstellung des Signals umfasst, und eines Maßes für eine Energie des Signals in dem Frequenzband;

Berechne (106) eines Maßes für eine Verteilung der Energie in dem Frequenzband, wobei die Verteilung der Energie in dem Frequenzband von einer vollständig gleichmäßigen Verteilung abweicht; und Berechne (104) des Schätzwerts unter Verwendung des Maßes für die Störung, des Maßes für die Energie und des Maßes für die Verteilung der Energie.

13. Computerprogramm mit einem Programmcode zum Durchführen des Verfahrens zum Ermitteln eines Schätzwerts für einen Bedarf an Informationseinheiten zum Codieren eines Signals gemäß Patentanspruch 12, wenn das Programm auf einem Computer abläuft.

Es folgen 7 Blatt Zeichnungen

Anhängende Zeichnungen

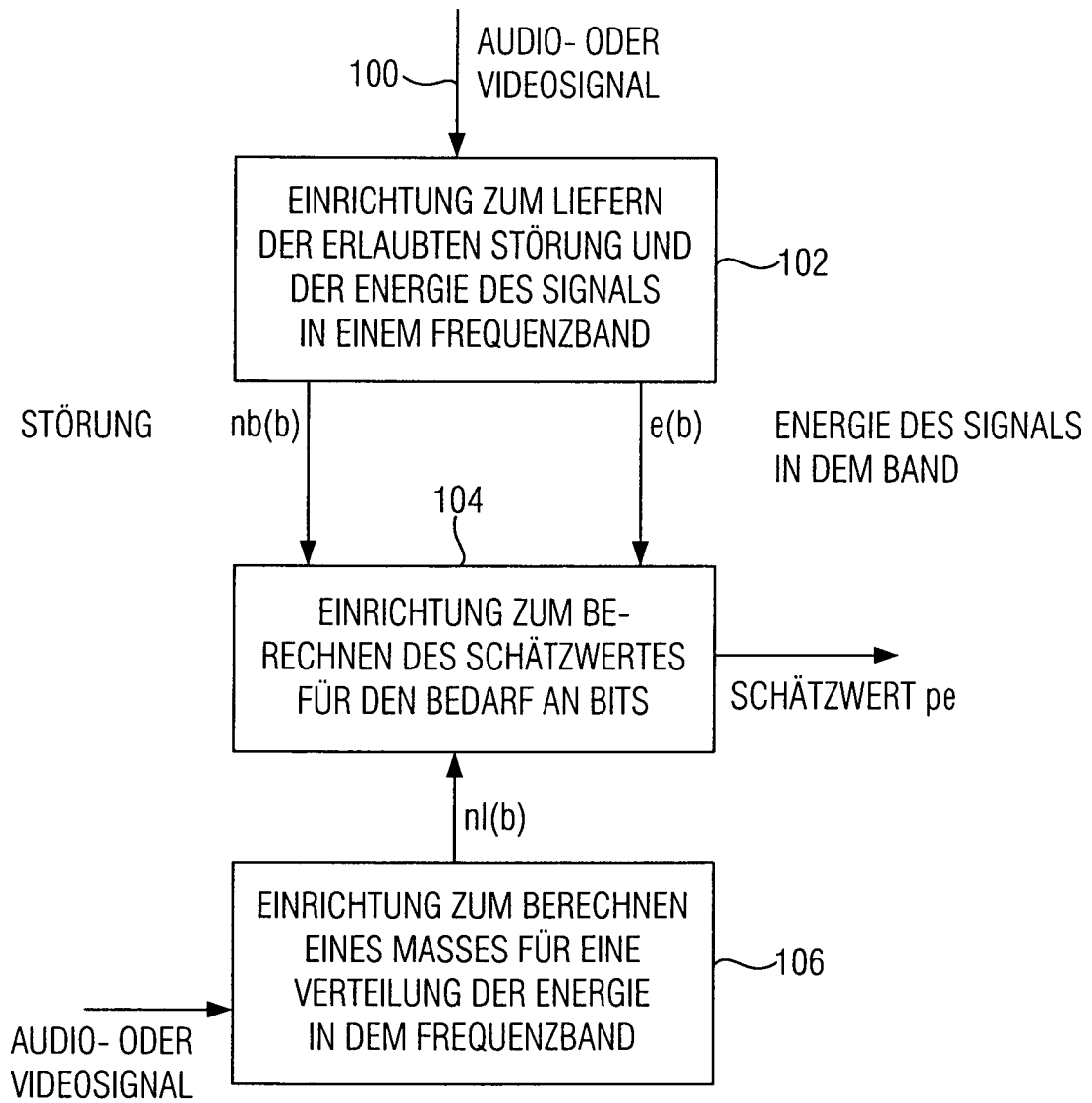


FIG. 1

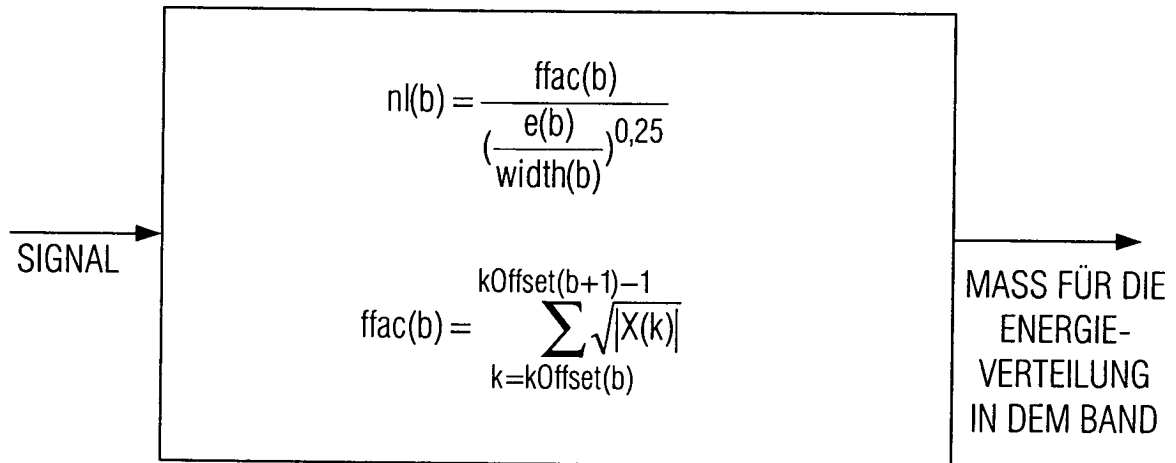


FIG. 2a

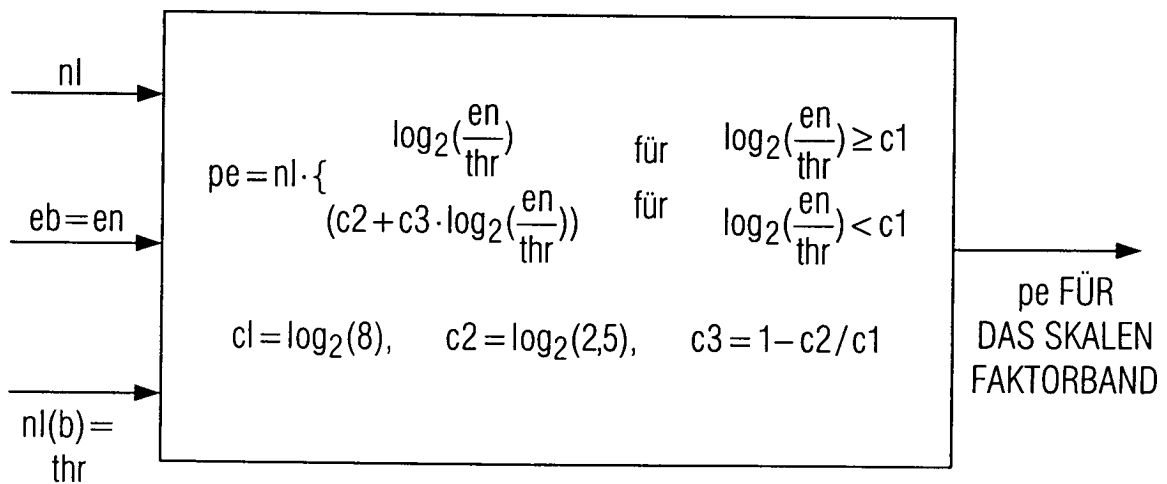


FIG. 2b

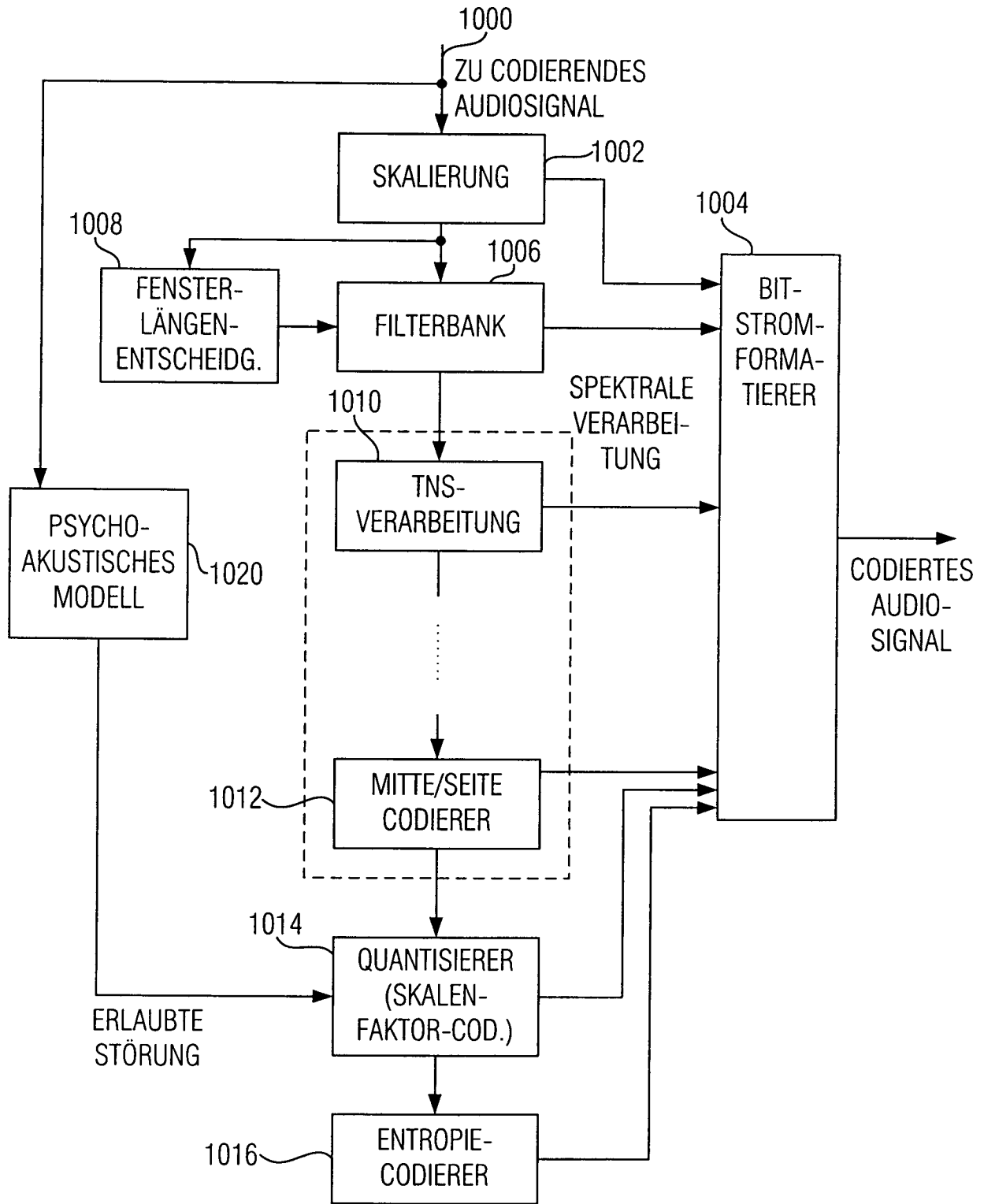


FIG. 3
(STAND DER TECHNIK)

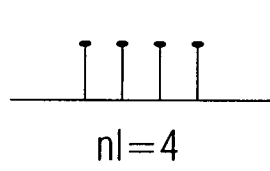


FIG. 4a

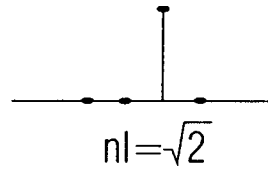


FIG. 4b

$$pe = nl(b) \cdot \log_2 \left(\frac{e(b)}{nb(b)} + 1,5 \right)$$

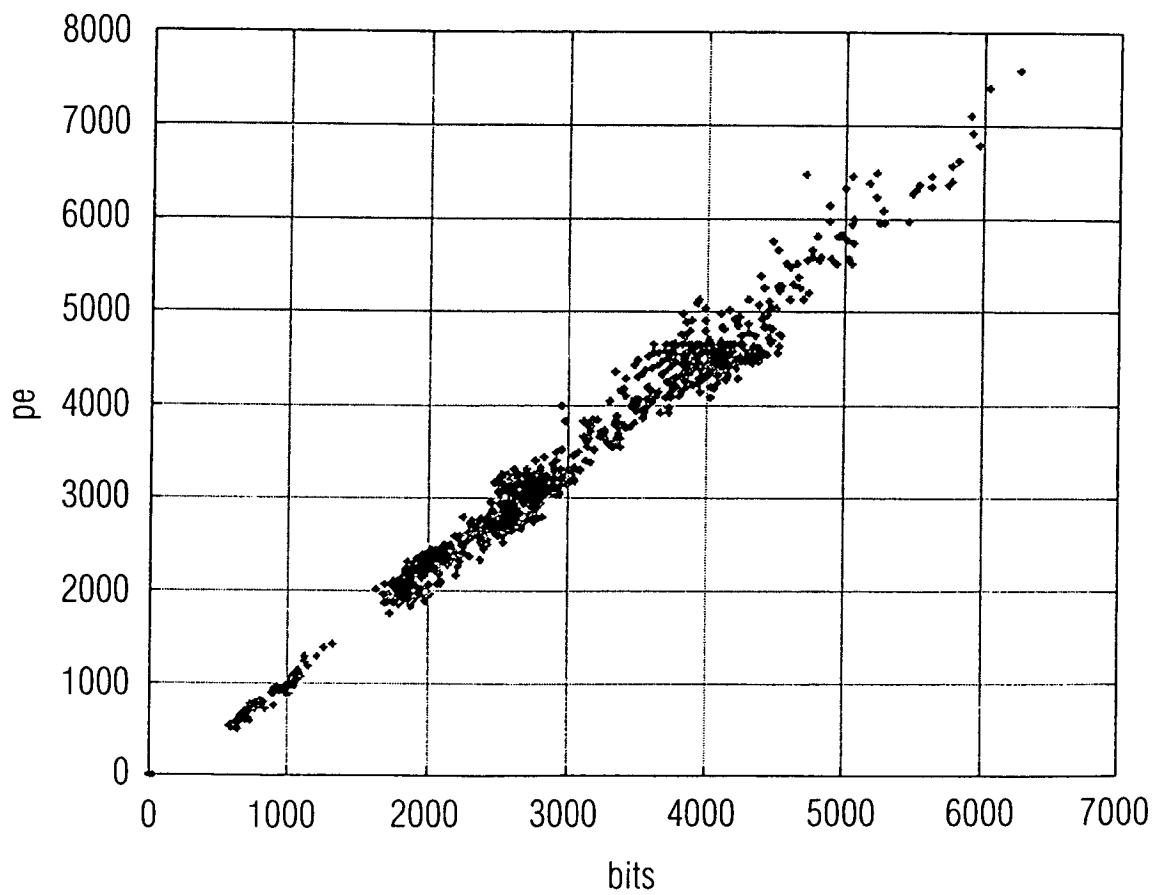
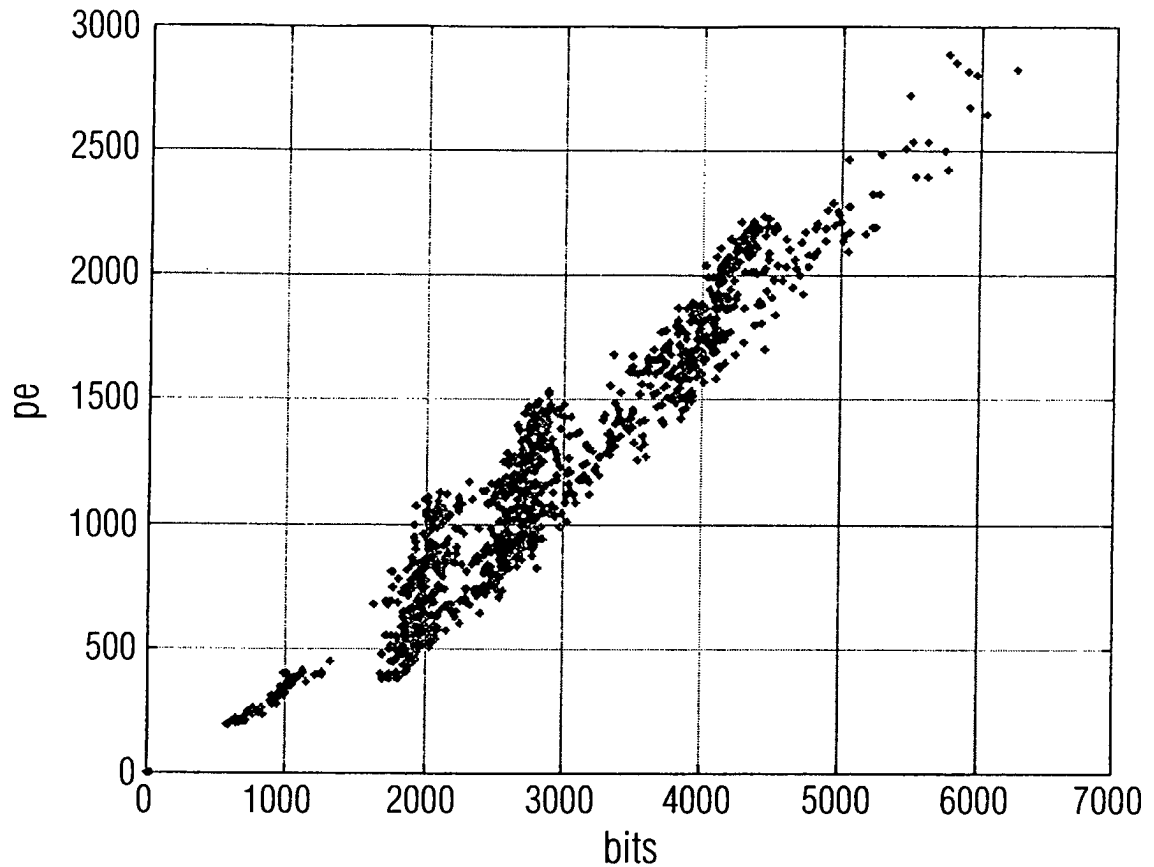


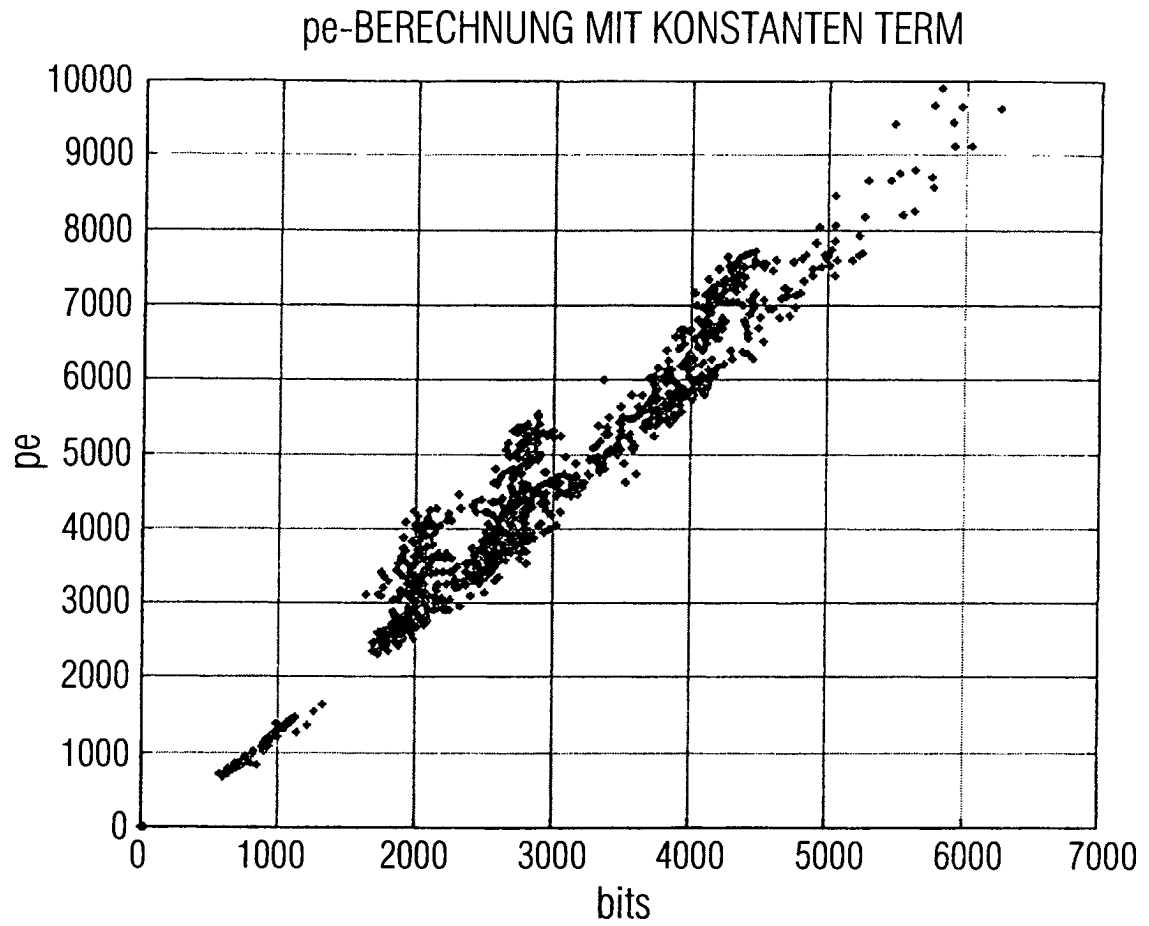
FIG. 5

pe BERECHNET NACH ISO/IEC IS 13818-7
MPEG -2 ADVANCED AUDIO CODING (AAC)



$$pe = -\sum_b \text{width}(b) \cdot \lg \left(\frac{nb(b)}{e(b)+1} \right)$$

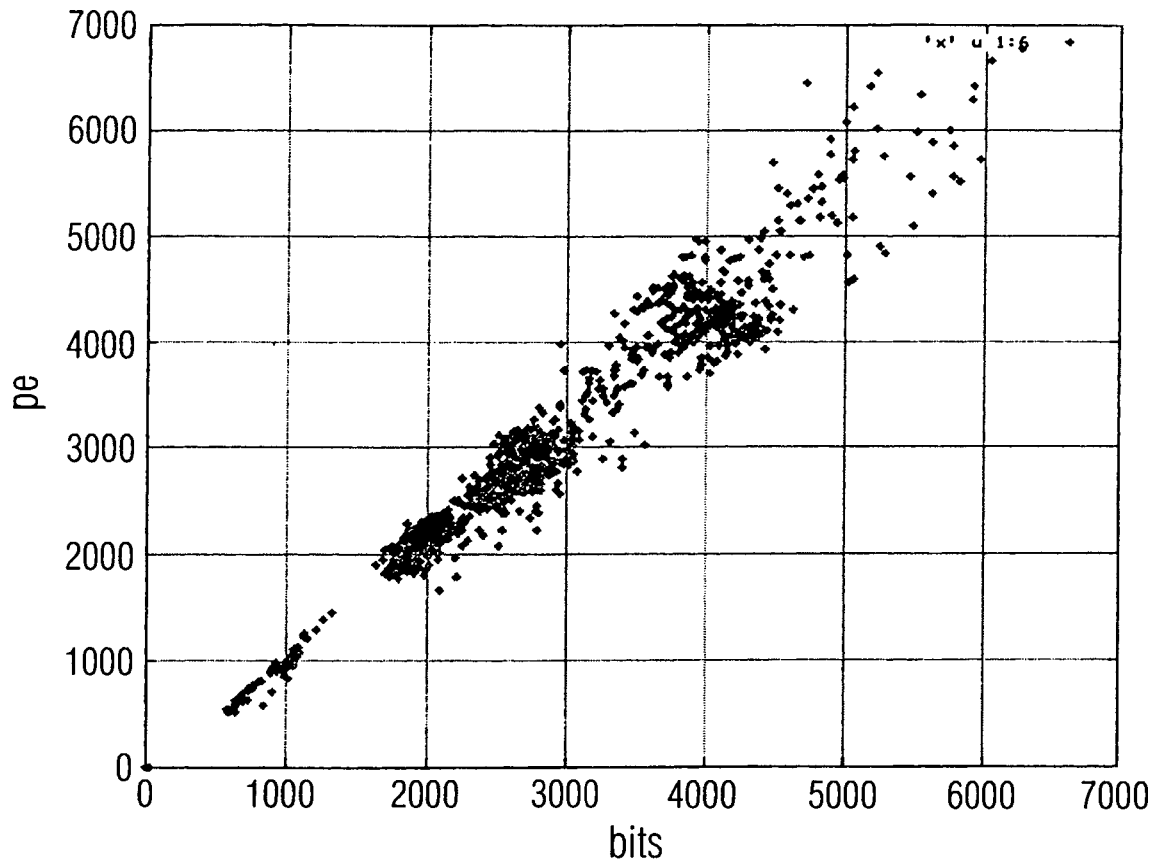
FIG. 6
(STAND DER TECHNIK)



$$pe = \sum_b \text{width}(b) \cdot \log_2 \left(\frac{e(b)}{nb(b)} + 1,5 \right)$$

FIG. 7

pe-BERECHNUNG LINIENWEISE MIT
KONSTANTEN TERM



$$pe = \sum_b \sum_{kOffset(b)}^{kOffset(b+1)-1} \log_2 \left(\frac{\chi^2(k) \cdot width(b)}{nb(b)} + 1,5 \right)$$

FIG. 8