

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 1 388 145 B1

(12)

EUROPÄISCHE PATENTSCHRIFT

(45) Veröffentlichungstag und Bekanntmachung des
Hinweises auf die Patenterteilung:
13.10.2004 Patentblatt 2004/42

(21) Anmeldenummer: **02745267.1**

(22) Anmeldetag: **25.04.2002**

(51) Int Cl.7: **G10L 11/00**

(86) Internationale Anmeldenummer:
PCT/EP2002/004618

(87) Internationale Veröffentlichungsnummer:
WO 2002/093557 (21.11.2002 Gazette 2002/47)

(54) **VORRICHTUNG UND VERFAHREN ZUM ANALYSIEREN EINES AUDIOSIGNALS HINSICHTLICH VON RHYTHMUSINFORMATIONEN**

DEVICE AND METHOD FOR ANALYSING AN AUDIO SIGNAL IN VIEW OF OBTAINING RHYTHM INFORMATION

DISPOSITIF ET PROCEDE POUR ANALYSER UN SIGNAL AUDIO AFIN D'OBTENIR DES INFORMATIONS DE RYTHME

(84) Benannte Vertragsstaaten:
AT CH DE FR GB LI

(30) Priorität: **14.05.2001 DE 10123366**

(43) Veröffentlichungstag der Anmeldung:
11.02.2004 Patentblatt 2004/07

(73) Patentinhaber: **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.**
80636 München (DE)

(72) Erfinder:
• **HERRE, Jürgen**
91054 Buckenhof (DE)
• **ROHDEN, Jan**
98693 Ilmenau (DE)
• **UHLE, Christian**
98693 Ilmenau (DE)
• **CREMER, Markus**
98693 Ilmenau (DE)

(74) Vertreter: **Schoppe, Fritz, Dipl.-Ing.**
Patentanwälte Schoppe, Zimmermann,
Stöckeler & Zinkler,
Postfach 246
82043 Pullach bei München (DE)

(56) Entgegenhaltungen:

- **SCHEIRER E D: "Tempo and beat analysis of acoustic musical signals" JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, JAN. 1998, ACOUST. SOC. AMERICA THROUGH AIP, USA, Bd. 103, Nr. 1, Seiten 588-601, XP002210418 ISSN: 0001-4966 in der Anmeldung erwähnt**
- **GOTO M ET AL: "Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, Bd. 27, Nr. 3-4, April 1999 (1999-04), Seiten 311-335, XP004163257 ISSN: 0167-6393**
- **SCHEIRER E D: "Pulse tracking with a pitch tracker" APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 1997. 1997 IEEE ASSP WORKSHOP ON NEW PALTZ, NY, USA 19-22 OCT. 1997, NEW YORK, NY, USA, IEEE, US, 19. Oktober 1997 (1997-10-19), Seite 4pp XP010248228 ISBN: 0-7803-3908-8 in der Anmeldung erwähnt**
- **TOLONEN T ET AL: "A computationally efficient multipitch analysis model" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, NOV. 2000, IEEE, USA, Bd. 8, Nr. 6, Seiten 708-716, XP002210419 ISSN: 1063-6676 in der Anmeldung erwähnt**

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).

EP 1 388 145 B1

Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf Signalverarbeitungskonzepte und insbesondere auf die Analyse von Audiosignalen hinsichtlich von Rhythmusinformationen.

[0002] In den letzten Jahren ist die Verfügbarkeit multimedialer Datenmaterials, wie z. B. Audio- oder Video-Daten, stark gestiegen. Dies ist auf eine Reihe von technischen Faktoren zurückzuführen, welche sich insbesondere auf die breite Verfügbarkeit des Internets, von leistungsfähiger Rechnerhard- und Software sowie von leistungsfähigen Verfahren zur Datenkompression, d. h. Quellcodierung, von Audio- und Videoverfahren stützen.

[0003] Die riesigen Mengen audiovisueller Daten, die beispielsweise auf dem Internet weltweit verfügbar sind, verlangen nach Konzepten, die es ermöglichen, diese Daten nach inhaltlichen Kriterien beurteilen, katalogisieren usw. zu können. Es besteht der Wunsch, in der Lage zu sein, multimediale Daten gezielt durch Angabe sinnvoller Kriterien zu suchen und aufzufinden.

[0004] Dies erfordert sogenannte "inhaltsbasierter" Techniken, die aus den audiovisuellen Daten sogenannte Merkmale, die in Fachkreisen auch als "Features" bezeichnet werden, extrahieren, welche wichtige charakteristische Eigenschaften des Signals repräsentieren. Basierend auf solchen Merkmalen bzw. Kombinationen dieser Merkmale können Ähnlichkeitsbeziehungen bzw. Gemeinsamkeiten zwischen Audio- oder Videosignalen hergeleitet werden. Dieser Vorgang erfolgt durch Vergleich bzw. In-Beziehung-Setzen der extrahierten Merkmalswerte aus den verschiedenen Signalen, welche auch einfach als "Stücke" bezeichnet werden.

[0005] Von besonderem Interesse ist die Bestimmung bzw. Extraktion von Merkmalen, die nicht nur signaltheoretische, sondern möglichst unmittelbare semantische Bedeutung haben, d. h. vom Hörer unmittelbar empfundene Eigenschaften repräsentieren.

[0006] Dies ermöglicht es dem Benutzer, auf einfache und intuitive Weise Suchanfragen zu formulieren, um Stücke aus dem gesamten vorhandenen Datenbestand einer Audiosignal-Datenbank zu finden. Ebenso erlauben es semantisch relevante Merkmale, Ähnlichkeitsbeziehungen zwischen Stücken zu modellieren, die der menschlichen Empfindung nahe kommen. Die Verwendung von Merkmalen, die semantische Bedeutung haben, ermöglicht auch beispielsweise ein automatisches Vorschlagen von für einen bestimmten Benutzer interessanten Stücken, wenn seine Vorlieben bekannt sind.

[0007] Im Bereich der Musikanalyse ist das Tempo ein wichtiger musikalischer Parameter, der semantische Bedeutung hat. Das Tempo wird üblicherweise in "Beats per Minute" (BPM) gemessen. Die automatische Extraktion des Tempos sowie der Taktschwerpunkte des "Beats" bzw. allgemein gesagt die automatische Extraktion von Rhythmusinformationen, ist ein Beispiel für die Gewinnung eines semantisch wichtigen Merkmals ei-

nes Musikstückes.

[0008] Weiterhin besteht ein Wunsch dahingehend, dass die Merkmalsextraktion, d. h. das Extrahieren von Rhythmusinformationen aus einem Audiosignal, robust und recheneffizient stattfinden kann. Robustheit bedeutet, dass es keine Rolle spielen darf, ob das Stück quellcodiert und wieder decodiert worden ist, ob das Stück über einen Lautsprecher abgespielt und von einem Mikrofon empfangen worden ist, ob es laut oder leise gespielt wird, oder ob es von einem Instrument oder einer Mehrzahl von Instrumenten gespielt wird.

[0009] Für die Bestimmung der Taktschwerpunkte und damit auch des Tempos, d. h. für die Bestimmung von Rhythmusinformationen, hat sich in den Fachkreisen auch der Begriff "Beat Tracking" etabliert. Aus dem Stand der Technik ist es bereits bekannt, ein Beat Tracking auf der Basis einer notenartigen bzw. transkribierten Signaldarstellung, z. B. im Midi-Format, durchzuführen. Ziel ist es jedoch, keine solchen Metadaten zu benötigen, sondern eine Analyse direkt mit einem z. B. PCMcodierten oder allgemein gesagt digital vorliegenden Audiosignal vorzunehmen.

[0010] Die Fachveröffentlichung "Tempo and Beat Analysis of Acoustic Musical Signals" von Eric D. Scheirer, J. Acoust. Soc. Am. 103:1, (Jan 1998), Seiten 588 - 601, offenbart ein Verfahren zur automatischen Extraktion eines rhythmischen Pulses aus musikalischen Exzerpten. Das Eingangssignal wird mittels einer Filterbank in eine Reihe von Teilbändern aufgespalten, beispielsweise in 6 Teilbänder mit Übergangsfrequenzen von 200 Hz, 400 Hz, 800 Hz, 1600 Hz und 3200 Hz. Für das erste Teilband wird eine Tiefpassfilterung durchgeführt. Für das letzte Teilband wird eine Hochpassfilterung durchgeführt, für die restlichen, dazwischen liegenden Teilbänder wird eine Bandpassfilterung beschrieben. Jedes Teilband wird folgendermaßen verarbeitet. Das Teilbandsignal wird zunächst gleichgerichtet. In anderen Worten ausgedrückt wird der Absolutbetrag der Abtastwerte bestimmt. Die resultierenden n Werte werden dann geglättet, beispielsweise mit einer Mittelung über ein geeignetes Fenster, um ein Hüllkurvensignal zu erhalten. Zur Senkung der Rechenkomplexität kann das Hüllkurvensignal unterabgetastet werden. Die Hüllkurvensignale werden differenziert, d. h. plötzliche Änderungen der Signalamplitude werden durch das Differenzierungsfilter bevorzugt weitergeleitet. Das Resultat wird dann auf nicht-negative Werte begrenzt. Jedes Hüllkurvensignal wird dann in eine Bank resonanter Filter, d. h. Oszillatoren, gegeben, die jeweils ein Filter für jeden Tempobereich enthalten, so dass das zum musikalischen Tempo passende Filter am stärksten angeregt wird. Für jedes Filter wird die Energie des Ausgangssignals als Maß für die Übereinstimmung des Tempos des Eingangssignals mit dem zum Filter gehörigen Tempo berechnet. Die Energien für jedes Tempo werden schließlich über alle Teilbänder aufsummiert, wobei die größte Energiesumme das als Resultat gelieferte Tempo, d. h. die Rhythmusinformationen, kenn-

zeichnet.

[0011] Ein wesentlicher Nachteil dieses Verfahrens besteht in der großen Rechen- und Speicherkomplexität insbesondere zur Realisierung der großen Zahl von parallelschwingenden "Oszillatoren", von denen letztlich nur ein einziger ausgewählt wird. Dies macht eine effiziente Implementierung beispielsweise für Echtzeitanwendungen nahezu unmöglich.

[0012] Die Fachveröffentlichung "Pulse Tracking with a Pitch Tracker" von Eric D. Scheirer, Proc. 1997 Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk, NY, Oct 1997, beschreibt einen Vergleich des oben beschriebenen "Oszillatorkonzeptes" mit einem Alternativkonzept, das auf der Verwendung von Autokorrelationsfunktionen zur Extraktion der Periodizität aus einem Audiosignal, d. h. der Rhythmusinformationen eines Signals, aufbaut. Ein Algorithmus zur Modulierung der menschlichen Tonhöhenwahrnehmung, d. h. des Pitch, wird für das "Beat Tracking" eingesetzt.

[0013] Der bekannte Algorithmus ist in Fig. 3 als Blockdiagramm dargestellt. Das Audiosignal wird über einen Audioeingang 300 einer Analyse-Filterbank 302 zugeführt. Die Analyse-Filterbank erzeugt aus dem Audioeingang eine Anzahl n von Kanälen, d. h. von einzelnen Teilbandsignalen. Jedes Teilbandsignal enthält einen bestimmten Bereich an Frequenzen des Audiosignals. Die Filter der Analyse-Filterbank sind so ausgewählt, dass sie die Selektionscharakteristik des menschlichen Innenohrs annähern. Eine solche Analyse-Filterbank wird auch als Gamma-Ton-Filterbank bezeichnet.

[0014] In den Einrichtungen 304a bis 304c werden die Rhythmus-Informationen jedes Teilbandsignals ausgewertet. Für jedes Eingangssignal wird zunächst ein hüllkurvenartiges Ausgangssignal berechnet (entsprechend einer sogenannten "Inner Hair Cell"-Verarbeitung im Ohr) und unterabgetastet. Aus diesem Resultat wird eine Autokorrelationsfunktion (AKF) berechnet, um die Periodizität des Signals als Funktion der Verzögerung, d. h. des "Lag" zu erhalten.

[0015] Am Ausgang der Einrichtungen 304a bis 304c liegt dann für jedes Teilbandsignal eine Autokorrelationsfunktion vor, welche Aspekte der Rhythmus-Informationen jedes Teilbandsignals darstellt.

[0016] Die einzelnen Autokorrelationsfunktionen der Teilbandsignale werden dann in einer Einrichtung 306 durch Summation kombiniert, um eine Summenautokorrelationsfunktion (SAKF) zu erhalten, welche die Rhythmus-Informationen des Signals am Audio-Eingang 300 wiedergibt. Diese Informationen können an einem Tempo-Ausgang 308 ausgegeben werden. Große Werte in der Summenautokorrelation zeigen an, dass für eine einer Spitze der SAKF zugeordneten Verzögerung (Lag) eine hohe Periodizität der Notenanfänge vorliegt. Daher wird beispielsweise der größte Wert der Summenautokorrelationsfunktion innerhalb der musikalisch sinnvollen Verzögerungen gesucht.

[0017] Musikalisch sinnvolle Verzögerungen sind beispielsweise der Tempobereich zwischen 60 bpm und 200 bpm. Die Einrichtung 306 kann ferner angeordnet sein, um eine Verzögerungszeit in Tempoinformationen umzusetzen. So entspricht beispielsweise eine Spitze einer Verzögerung von einer Sekunde einem Tempo von 60 Schlägen pro Minute. Kleinere Verzögerungen deuten auf höhere Tempos hin, während größere Verzögerungen auf kleinere Tempos als 60 bpm hinweisen.

[0018] Dieses Verfahren hat gegenüber dem zuerst genannten Verfahren einen Vorteil dahingehend, dass keine Oszillatoren mit großem Rechen- und Speicheraufwand implementiert werden müssen. Andererseits ist das Konzept dahingehend nachteilhaft, dass die Qualität der Ergebnisse sehr stark von der Art des Audiosignals abhängt. Ist aus einem Audiosignal beispielsweise ein dominantes Rhythmusinstrument herauszuhören, so wird das in Fig. 3 beschriebene Konzept gut funktionieren. Ist dagegen die Stimme dominant, welche keine besonders eindeutigen Rhythmusinformationen liefern wird, so wird die Rhythmusbestimmung mehrdeutig sein. In dem Audiosignal könnte durchaus auch ein Band vorkommen, das lediglich Rhythmusinformationen enthält, wie z. B. ein höheres Frequenzband, in dem beispielsweise ein Hi-hat eines Schlagzeugs positioniert ist, oder ein niedriges Frequenzband, in dem die große Trommel eines Schlagzeugs auf der Frequenzskala positioniert ist. Aufgrund der Kombination der Einzelinformationen werden jedoch die einigermaßen eindeutigen Informationen dieser speziellen Teilbänder von den mehrdeutigen Informationen der anderen Teilbänder überlagert bzw. "verwässert".

[0019] Ein weiteres Problem bei der Verwendung von Autokorrelationsfunktionen zum Extrahieren der Periodizität eines Teilbandsignals besteht darin, dass die Summen-Autokorrelationsfunktion, die durch die Einrichtung 306 gewonnen wird, mehrdeutig ist. Die Summen-Autokorrelationsfunktion am Ausgang 306 ist dahingehend mehrdeutig, dass auch beim Vielfachen einer Verzögerung eine Autokorrelationsfunktions-Spitze erzeugt wird. Dies ist daraus verständlich, dass eine Sinuskomponente mit einer Periode von t_0 , wenn sie einer Autokorrelationsfunktionverarbeitung unterzogen wird, neben dem gewünschten Maximum bei t_0 auch Maxima bei Vielfachen der Verzögerungen, d. h. bei $2t_0$, $3t_0$, usw. erzeugt.

[0020] Die Fachveröffentlichung "A Computationally Efficient Multipitch Analysis Model", von Tolonen und Karjalainen, IEEE Transactions on Speech and Audio Processing, Band 8, Nr. 6, Nov. 2000, offenbart ein rechenzeiteffizientes Modell für eine Periodizitätsanalyse von komplexen Audiosignalen. Das Rechenmodell teilt das Signal in zwei Kanäle auf, und zwar in einen Kanal unter 1000 Hz und einen Kanal über 1000 Hz. Hieraus wird eine Autokorrelation des unteren Kanals und eine Autokorrelation der Hüllkurve des oberen Kanals berechnet. Schließlich werden die beiden Autokorrelationsfunktionen summiert. Um die Mehrdeutigkeiten der

Summenautokorrelationsfunktion zu eliminieren, wird die Summenautokorrelationsfunktion weiterverarbeitet, um eine sogenannte Enhanced Summary Autocorrelation Function (ESACF) (Weiterentwickelte Summenautokorrelationsfunktion) zu erhalten. Diese Nachbearbeitung der Summenautokorrelationsfunktion beinhaltet ein wiederholtes Abziehen von mit ganzzahligen Faktoren gespreizten Versionen der Autokorrelationsfunktion von der Summenautokorrelationsfunktion mit nachfolgender Begrenzung auf nicht-negative Werte.

[0021] Die Aufgabe der vorliegenden Erfindung besteht darin, eine rechenzeiteffiziente und robuste Vorrichtung und ein rechenzeiteffizientes und robustes Verfahren zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen zu schaffen.

[0022] Diese Aufgabe wird durch eine Vorrichtung zum Analysieren eines Audiosignals nach Patentanspruch 1 oder durch ein Verfahren zum Analysieren eines Audiosignals nach Patentanspruch 11 gelöst.

[0023] Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, dass in den einzelnen Frequenzbändern, d. h. den Teilbändern, oftmals unterschiedlich günstige Bedingungen für das Auffinden von rhythmischen Periodizitäten herrschen. Während beispielsweise bei Popmusik oftmals im Bereich der Mitte, beispielsweise um 1 kHz, das Signal von nicht mit dem Beat korrespondierenden Gesang dominiert wird, sind in den höheren Frequenzbereichen oftmals vor allem Perkussionsklänge präsent, wie z. B. das Hi-hat des Schlagzeugs, welche eine sehr gute Extraktion rhythmischer Regelmäßigkeiten erlauben. Anders ausgedrückt beinhalten unterschiedliche Frequenzbänder je nach Audiosignal eine unterschiedliche Menge an rhythmischer Information bzw. haben eine unterschiedliche Qualität oder Signifikanz für die Rhythmusinformationen des Audiosignals.

[0024] Erfindungsgemäß wird daher das Audiosignal zunächst in Teilbandsignale zerlegt. Jedes Teilbandsignal wird hinsichtlich seiner Periodizität untersucht, um Rhythmus-Rohinformationen für jedes Teilbandsignal zu erhalten. Hierauf wird gemäß der vorliegenden Erfindung eine Bewertung der Qualität der Periodizität jedes Teilbandsignals durchgeführt, um ein Signifikanzmaß für jedes Teilbandsignal zu erhalten. Ein hohes Signifikanzmaß deutet darauf hin, dass in diesem Teilbandsignal deutliche Rhythmusinformationen vorliegen, während ein niedriges Signifikanzmaß darauf hindeutet, dass in diesem Teilbandsignal weniger eindeutige Rhythmusinformationen vorliegen.

[0025] Gemäß einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird bei der Untersuchung eines Teilbandsignals hinsichtlich seiner Periodizitäten zunächst eine modifizierte Hüllkurve des Teilbandsignals berechnet und dann eine Autokorrelationsfunktion der Hüllkurve berechnet. Die Autokorrelationsfunktion der Hüllkurve stellt die Rhythmus-Rohinformationen dar. Eindeutige Rhythmusinformationen liegen vor, wenn die Autokorrelationsfunktion deutliche Maxima aufweist, während weniger eindeutige Rhythmusin-

formationen vorliegen, wenn die Autokorrelationsfunktion der Hüllkurve des Teilbandsignals weniger ausgeprägte Signalspitzen oder überhaupt keine Signalspitzen hat. Eine Autokorrelationsfunktion, die deutliche Signalspitzen hat, wird daher ein hohes Signifikanzmaß erhalten, während eine Autokorrelationsfunktion, die einen relativ flachen Verlauf hat, ein niedriges Signifikanzmaß erhalten wird.

[0026] Erfindungsgemäß werden die einzelnen Rhythmus-Rohinformationen der einzelnen Teilbandsignale nicht einfach "blind" kombiniert, sondern unter Berücksichtigung des Signifikanzmaßes für jedes Teilbandsignal verwendet, um die Rhythmusinformationen des Audiosignals zu erhalten. Hat ein Teilbandsignal ein hohes Signifikanzmaß, so wird es bei der Ermittlung der Rhythmusinformationen bevorzugt, während ein Teilbandsignal, das ein niedriges Signifikanzmaß hat, d. h. das eine niedrige Qualität hinsichtlich der Rhythmusinformationen aufweist, bei der Ermittlung der Rhythmusinformationen des Audiosignals kaum oder im Extremfall überhaupt nicht berücksichtigt wird.

[0027] Dies kann rechenzeiteffizient gut durch einen Gewichtungsfaktor implementiert werden, der von dem Signifikanzmaß abhängt. Während ein Teilbandsignal, das eine gute Qualität für die Rhythmusinformationen hat, d. h. das ein hohes Signifikanzmaß hat, einen Gewichtungsfaktor von 1 erhalten könnte, wird ein anderes Teilbandsignal, das ein kleineres Signifikanzmaß hat, einen Gewichtungsfaktor kleiner als 1 erhalten. Im Extremfall wird ein Teilbandsignal, das eine vollkommen flache Autokorrelationsfunktion hat, einen Gewichtungsfaktor von 0 haben. Die gewichteten Autokorrelationsfunktionen, d. h. die gewichteten Rhythmus-Rohinformationen werden dann einfach aufsummiert. Wenn lediglich ein Teilbandsignal sämtlicher Teilbandsignale gute Rhythmusinformationen liefert, während die anderen Teilbandsignale Autokorrelationsfunktionen mit einem flachen Verlauf aufweisen, kann diese Gewichtung im Extremfall dazu führen, dass alle Teilbandsignale außer dem einen Teilbandsignal einen Gewichtungsfaktor von 0 erhalten, d. h. bei der Ermittlung der Rhythmusinformationen überhaupt nicht berücksichtigt werden, so dass die Rhythmusinformationen des Audiosignals lediglich aus einem einzigen Teilbandsignal ermittelt werden.

[0028] Das erfindungsgemäße Konzept ist dahingehend vorteilhaft, dass es eine robuste Bestimmung der Rhythmusinformationen ermöglicht, da Teilbandsignale mit keinen eindeutigen bzw. sogar abweichenden Rhythmusinformationen, d. h. wenn der Gesang einen anderen Rhythmus aufweist als der eigentliche Beat des Stücks, die Rhythmusinformationen des Audiosignals nicht "verwässern" bzw. "verfälschen". Darüber hinaus werden sehr rauschartige Teilbandsignale, welche eine Systemautokorrelationsfunktion mit vollständig flachem Verlauf liefern, das Signal/Rausch-Verhältnis bei der Bestimmung der Rhythmusinformationen nicht verschlechtern. Genau dies würde jedoch auftre-

ten, wenn, wie im Stand der Technik, einfach alle Autokorrelationsfunktionen der Teilbandsignale mit gleichem Gewicht aufsummiert werden.

[0029] Ein weiterer Vorteil des erfindungsgemäßen Verfahrens besteht darin, dass mit einem kleinen zusätzlichen Rechenaufwand ein Signifikanzmaß bestimmt werden kann, und dass die Bewertung der Rhythmus-Rohinformationen mit dem Signifikanzmaß und die anschließende Summierung ohne großen Speicher- und Rechenzeitaufwand effizient durchgeführt werden können, was das erfindungsgemäße Konzept insbesondere auch für Echtzeitanwendungen empfiehlt.

[0030] Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend bezugnehmend auf die beiliegenden Zeichnungen detailliert erläutert. Es zeigen:

Fig. 1 ein Blockschaltbild einer Vorrichtung zum Analysieren eines Audiosignals mit einer Qualitätsbewertung der Rhythmus-Rohinformationen;

Fig. 2 ein Blockschaltbild einer Vorrichtung zum Analysieren eines Audiosignals unter Verwendung von Gewichtungsfaktoren auf der Basis der Signifikanzmaße;

Fig. 3 ein Blockschaltbild einer bekannten Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen;

Fig. 4 ein Blockschaltbild einer Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen unter Verwendung einer Autokorrelationsfunktion mit einer teilbandweisen Nachbearbeitung der Rhythmus-Rohinformationen; und

Fig. 5 ein detailliertes Blockschaltbild der Einrichtung zum Nachbearbeiten von Fig. 4.

[0031] Fig. 1 zeigt ein Blockschaltbild einer Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen. Das Audiosignal wird über einen Eingang 100 einer Einrichtung 102 zum Zerlegen des Audiosignals in zumindest zwei Teilbandsignale 104a und 104b zugeführt. Jedes Teilbandsignal 104a, 104b wird in eine Einrichtung 106a bzw. 106b zum Untersuchen desselben hinsichtlich von Periodizitäten in dem Teilbandsignal zugeführt, um Rhythmus-Rohinformationen 108a bzw. 108b für jedes Teilbandsignal zu erhalten. Die Rhythmus-Rohinformationen werden dann einer Einrichtung 110a bzw. 110b zum Bewerten einer Qualität der Periodizität jedes der zumindest zwei Teilbandsignale zugeführt, um ein Signifikanzmaß 112a, 112b für jedes der zumindest zwei Teilbandsignale zu erhalten. Sowohl die Rhythmus-Rohinformationen

108a, 108b als auch die Signifikanzmaße 112a, 112b werden einer Einrichtung 114 zum Ermitteln der Rhythmusinformationen des Audiosignals zugeführt. Die Einrichtung 114 berücksichtigt beim Ermitteln der Rhythmusinformationen des Audiosignals die Signifikanzmaße 112a, 112b für die Teilbandsignale sowie die Rhythmus-Rohinformationen 108a, 108b von zumindest einem Teilbandsignal.

[0032] Hat die Einrichtung 110a zur Qualitätsbewertung beispielsweise festgestellt, dass in dem Teilbandsignal 104a keine besondere Periodizität vorliegt, so wird das Signifikanzmaß 112a sehr klein, bzw. gleich 0 sein. In diesem Fall stellt die Einrichtung 114 zum Ermitteln der Rhythmusinformationen fest, dass das Signifikanzmaß 112a gleich Null ist, so dass die Rhythmus-Rohinformationen 108a des Teilbandsignals 104a bei der Ermittlung der Rhythmusinformationen des Audiosignals überhaupt nicht mehr berücksichtigt werden müssen. Die Rhythmusinformationen des Audiosignals werden dann allein und ausschließlich auf der Basis der Rhythmus-Rohinformationen 108b des Teilbandsignals 104b bestimmt.

[0033] Im nachfolgenden wird auf Fig. 2 hinsichtlich einer speziellen Ausführungsform der Vorrichtung von Fig. 1 eingegangen. Als Einrichtung 102 zum Zerlegen des Audiosignals kann eine übliche Analyse-Filterbank verwendet werden, die ausgangsseitig eine von einem Benutzer wählbare Anzahl von Teilbandsignalen liefert. Jedes Teilbandsignal wird dann der Verarbeitung der Einrichtungen 106a, 106b bzw. 106c unterzogen, woraufhin dann durch die Einrichtungen 110a bis 110c von jeder Rhythmus-Rohinformationen Signifikanzmaße ermittelt werden. Die Einrichtung 114 umfaßt bei der in Fig. 2 dargestellten bevorzugten Ausführungsform eine Einrichtung 114a zum Berechnen von Gewichtungsfaktoren für jedes Teilbandsignal auf der Basis des Signifikanzmaßes für dieses Teilbandsignal und optional auch der anderen Teilbandsignale. In der Einrichtung 114b findet dann eine Gewichtung der Rhythmus-Rohinformationen 108a bis 108c mit dem Gewichtungsfaktor für dieses Teilbandsignal statt, woraufhin dann, ebenfalls in der Einrichtung 114b, die gewichteten Rhythmus-Rohinformationen kombiniert, z. B. aufsummiert, werden, um an dem Tempo-Ausgang 116 die Rhythmusinformationen des Audiosignals zu erhalten.

[0034] Das erfindungsgemäße Konzept stellt sich somit folgendermaßen dar. Nach der Evaluierung der rhythmischen Informationen der Einzelbänder, welche beispielsweise durch Hüllkurvenbildung, Glättung, Differenzieren, Begrenzen auf positive Werte und Bilden der Autokorrelationsfunktion stattfinden kann (Einrichtungen 106a bis 106c), findet eine Bewertung der Wertigkeit bzw. der Qualität dieser Zwischenergebnisse in den Einrichtungen 110a bis 110c statt. Dies wird mit Hilfe einer Bewertungsfunktion erreicht, welche die Verlässlichkeit der jeweiligen Einzelergebnisse mit einem Signifikanzmaß bewertet. Aus den Signifikanzmaßen aller Teilbandsignale wird ein Gewichtungsfaktor für jedes

Band für die Extraktion der Rhythmus-Informationen hergeleitet. Das Gesamtergebnis der Rhythmusextraktion wird dann in der Einrichtung 114b durch Kombination der bandweisen Einzelergebnisse unter Berücksichtigung ihrer jeweiligen Gewichtungsfaktoren erreicht.

[0035] Als Resultat zeigt ein solchermaßen implementierter Algorithmus zur Rhythmusanalyse eine gute Fähigkeit, rhythmische Informationen in einem Signal auch unter ungünstigen Bedingungen zuverlässig zu finden. Das erfindungsgemäße Konzept zeichnet sich daher durch eine hohe Robustheit aus.

[0036] Bei einer bevorzugten Ausführungsform werden die Rhythmus-Rohinformationen 108a, 108b, 108c, die die Periodizität des jeweiligen Teilbandsignals darstellen, mittels einer Autokorrelationsfunktion bestimmt. In diesem Fall wird es bevorzugt, das Signifikanzmaß zu bestimmen, indem ein Maximum der Autokorrelationsfunktion durch einen Mittelwert der Autokorrelationsfunktion geteilt wird, und dann der Wert 1 subtrahiert wird. Es sei darauf hingewiesen, daß jede Autokorrelationsfunktion immer bei einer Verzögerung von 0 ein lokales Maximum liefert, das die Energie des Signals repräsentiert. Dieses Maximum sollte unberücksichtigt bleiben, damit die Qualitätsbestimmung nicht verfälscht wird.

[0037] Ferner soll die Autokorrelationsfunktion lediglich in einem speziellen Tempobereich betrachtet werden, d. h. von einer maximalen Verzögerung, die dem kleinsten interessierenden Tempo entspricht, zu einer minimalen Verzögerung, die dem höchsten interessierenden Tempo entspricht. Ein typischer Tempobereich liegt zwischen 60 bpm und 200 bpm.

[0038] Alternativ kann als Signifikanzmaß das Verhältnis zwischen dem arithmetischen Mittelwert der Autokorrelationsfunktion im interessierenden Tempobereich und dem geometrischen Mittelwert der Autokorrelationsfunktion im interessierenden Tempobereich bestimmt werden. Es ist bekannt, dass wenn alle Werte der Autokorrelationsfunktion gleich sind, d. h. wenn die Autokorrelationsfunktion einen flachen Verlauf hat, der geometrische Mittelwert der Autokorrelationsfunktion und der arithmetische Mittelwert der Autokorrelationsfunktion gleich sind. In diesem Fall hätte das Signifikanzmaß einen Wert gleich 1, was bedeutet, dass die Rhythmus-Rohinformationen nicht signifikant sind.

[0039] Im Falle einer Systemautokorrelationsfunktion mit starken Spitzen würde das Verhältnis aus arithmetischem Mittelwert zu geometrischem Mittelwert größer 1 sein, was bedeutet, dass die Autokorrelationsfunktion gute Rhythmusinformationen aufweist. Je kleiner jedoch das Verhältnis zwischen arithmetischem Mittelwert und geometrischem Mittelwert wird, desto flacher ist die Autokorrelationsfunktion und desto weniger Periodizitäten enthält sie, was wiederum bedeutet, dass die Rhythmusinformationen dieses Teilbandsignals weniger signifikant, d. h. eine geringere Qualität haben, was sich in einem geringen bzw. einem Gewichtungsfaktor von 0 äußern wird.

faktor von 0 äußern wird.

[0040] Bezüglich der Gewichtungsfaktoren existieren verschiedene Möglichkeiten. Bevorzugt wird eine relative Gewichtung, derart, dass sich alle Gewichtungsfaktoren sämtlicher Teilbandsignale zu 1 aufaddieren, d. h. dass der Gewichtungsfaktor eines Bandes bestimmt wird als der Signifikanzwert dieses Bandes geteilt durch die Summe aller Signifikanzwerte. In diesem Fall wird eine relative Gewichtung vor der Aufsummation der gewichteten Rhythmus-Rohinformationen durchgeführt, um die Rhythmus-Informationen des Audiosignals zu erhalten.

[0041] Wie bereits ausgeführt worden ist, wird es bevorzugt, die Auswertung der Rhythmus-Informationen unter Verwendung einer Autokorrelationsfunktion durchzuführen. Dieser Fall ist in Fig. 4 dargestellt. Das Audiosignal wird über den Audiosignaleingang 100 in die Einrichtung 102 zum Zerlegen des Audiosignals in Teilbandsignale 104a und 104b eingespeist. Jedes Teilbandsignal wird dann in der Einrichtung 106a bzw. 106b, wie es ausgeführt worden ist, unter Verwendung einer Autokorrelationsfunktion untersucht, um die Periodizität des Teilbandsignals zu ermitteln. Am Ausgang der Einrichtung 106a bzw. 106b liegen dann die Rhythmus-Rohinformationen 108a, 108b vor. Diese werden in eine Einrichtung 118a bzw. 118b eingespeist, um die mittels der Autokorrelationsfunktion von der Einrichtung 116a ausgegebenen Rhythmus-Rohinformationen nachzubearbeiten. Damit wird u. a. sichergestellt, daß die Mehrdeutigkeiten der Autokorrelationsfunktion, d. h. daß bei ganzzahligen Vielfachen der Verzögerungen ebenfalls Signalspitzen auftreten, teilbandweise eliminiert werden, um nachbearbeitete Rhythmus-Rohinformationen 120a bzw. 120b zu erhalten.

[0042] Dies hat den Vorteil, dass die Mehrdeutigkeiten der Autokorrelationsfunktionen, d. h. der Rhythmus-Rohinformationen 108a, 108b, bereits teilbandweise eliminiert werden, und nicht erst, wie im Stand der Technik, nach der Summation der einzelnen Autokorrelationsfunktionen. Darüber hinaus ermöglicht die einzelbandweise Eliminierung der Mehrdeutigkeiten in den Autokorrelationsfunktionen durch die Einrichtungen 118a, 118b, dass die Rhythmus-Rohinformationen der Teilbandsignale unabhängig voneinander gehandhabt werden können. Sie können beispielsweise einer Qualitätsbeurteilung mittels der Einrichtung 110a für die Rhythmus-Rohinformationen 108a oder mittels der Einrichtung 110b für die Rhythmus-Rohinformationen 108b unterzogen werden.

[0043] Wie es durch die gestrichelten Linien in Fig. 4 dargestellt ist, kann die Qualitätsbeurteilung jedoch auch anhand der nachbearbeiteten Rhythmus-Rohinformationen stattfinden, wobei diese letztere Möglichkeit bevorzugt wird, da die Qualitätsbeurteilung auf der Basis der nachbearbeiteten Rhythmus-Rohinformationen sicherstellt, dass die Qualität einer Information beurteilt wird, die nicht mehr vieldeutig ist.

[0044] Die Ermittlung der Rhythmus-Informationen

durch die Einrichtung 114 findet dann auf der Basis von nachbearbeiteten Rhythmus-Informationen eines Kanals und vorzugsweise auch auf der Basis des Signifikanzmaßes für diesen Kanal statt.

[0045] Wenn eine Qualitätsbeurteilung auf der Basis der Rhythmus-Rohinformationen, also des Signals vor der Einrichtung 118a durchgeführt wird, ist dies dahingehend vorteilhaft, dass, wenn bestimmt wird, dass das Signifikanzmaß gleich 0 ist, d. h. dass die Autokorrelationsfunktion einen flachen Verlauf hat, auf die Nachbearbeitung mittels der Einrichtung 118a gänzlich verzichtet werden kann, um Rechenzeitressourcen einzusparen.

[0046] Im nachfolgenden wird auf Fig. 5 eingegangen, um einen detaillierteren Aufbau einer Einrichtung 118a oder 118b zur Nachbearbeitung der Rhythmus-Rohinformationen darzustellen. Zunächst wird das Teilband-Signal beispielsweise 104a, in die Einrichtung 106a zur Untersuchung der Periodizität des Teilbandsignals mittels einer Autokorrelationsfunktion eingepeist, um Rhythmus-Rohinformationen 108a zu erhalten. Um die Mehrdeutigkeiten teilbandweise zu eliminieren, kann genauso wie im Stand der Technik, eine gespreizte Autokorrelationsfunktion mittels einer Einrichtung 121 berechnet werden, wobei die Einrichtung 121 angeordnet ist, um die gespreizte Autokorrelationsfunktion so zu berechnen, dass sie um ein ganzzahliges Vielfaches einer Verzögerung gespreizt ist. Eine Einrichtung 122 ist in diesem Fall angeordnet, um die gespreizte Autokorrelationsfunktion von der ursprünglichen Autokorrelationsfunktion, d. h. den Rhythmus-Rohinformationen 108a zu subtrahieren. Insbesondere wird es bevorzugt, zunächst eine auf das Doppelte gespreizte Autokorrelationsfunktion in der Einrichtung 121 zu berechnen und dann von den Rhythmus-Rohinformationen 108a zu subtrahieren. Dann, in dem nächsten Schritt, wird eine um den Faktor 3 gespreizte Autokorrelationsfunktion in der Einrichtung 121 berechnet und von dem Ergebnis der vorherigen Subtraktion wieder abgezogen, so dass nach und nach sämtliche Mehrdeutigkeiten aus den Rhythmus-Rohinformationen eliminiert werden.

[0047] Alternativ oder zusätzlich kann die Einrichtung 121 angeordnet sein, um eine um einen ganzzahligen Faktor gestauchte Autokorrelationsfunktion zu berechnen, wobei diese dann von der Einrichtung 122 zu den Rhythmus-Rohinformationen hinzu addiert wird, um auch Anteile für Verzögerungen $t_0/2$, $t_0/3$ usw. zu erzeugen.

[0048] Darüber hinaus können die gespreizten bzw. gestauchten Versionen der Rhythmus-Rohinformationen 108a vor dem Addieren bzw. Subtrahieren gewichtet werden, um auch hier eine Flexibilität im Sinne einer hohen Robustheit zu erreichen.

[0049] Durch das Verfahren, die Periodizität eines Teilbandsignals auf der Basis einer Autokorrelationsfunktion zu untersuchen, kann also eine weitere Verbesserung erzielt werden wenn die Eigenschaften der Au-

tokorrelationsfunktion mit einbezogen werden und die Nachbearbeitung unter Verwendung der Einrichtung 118a oder 118b durchgeführt wird. So erzeugt eine periodische Folge von Notenanfängen mit einem Abstand t_0 nicht nur eine AKF-Spitze bei einer Verzögerung t_0 sondern auch bei $2t_0$, $3t_0$, etc. Dies wird zu einer Vieldeutigkeit in der Tempodetektion, d. h. dem Suchen signifikanter Maxima in der Autokorrelationsfunktion, führen. Die Mehrdeutigkeiten können dadurch eliminiert werden, wenn um ganzzahlige Faktoren gespreizte Versionen der AKF vom Ausgangswert teilbandweise (gewichtet) abgezogen werden.

[0050] Ferner besteht bei der Autokorrelationsfunktion das Problem, daß sie keine Informationen bei $t_0/2$, $t_0/3$... usw., also beim Doppelten, Dreifachen, etc. des "Grundtempos" liefert, was besonders dann zu falschen Ergebnissen führen kann, wenn zwei Instrumente, die in unterschiedlichen Teilbändern liegen, zusammen den Rhythmus des Signals definieren. Diese Sache wird dadurch berücksichtigt, dass um ganzzahlige Faktoren gestauchte Versionen der Autokorrelationsfunktion berechnet werden und diese dann zu den Rhythmus-Rohinformationen gewichtet oder ungewichtet hinzugeaddiert werden.

[0051] Die AKF-Nachverarbeitung findet somit teilbandweise statt, wobei für mindestens ein Teilbandsignal eine Autokorrelationsfunktion errechnet wird und diese mit gedehnten oder gespreizten Versionen dieser Funktion kombiniert wird.

Patentansprüche

1. Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen des Audiosignals, mit folgenden Merkmalen:

einer Einrichtung (102) zum Zerlegen des Audiosignals in zumindest zwei Teilbandsignale (104a, 104b);

einer Einrichtung zum Untersuchen (106a, 106b) eines Teilbandsignals hinsichtlich einer Periodizität in dem Teilbandsignal, um Rhythmus-Rohinformationen (108a, 108b) für das Teilbandsignal zu erhalten;

einer Einrichtung zum Bewerten (110a, 110b) einer Qualität der Periodizität der Rhythmus-Rohinformationen (108a) des Teilbandsignals (104a), um ein Signifikanzmaß (112a) für das Teilbandsignal zu erhalten; und

einer Einrichtung (114) zum Ermitteln der Rhythmusinformationen des Audiosignals unter Berücksichtigung des Signifikanzmaßes (112a) des Teilbandsignals und der Rhythmus-Rohinformationen (108a, 108b) von zumindest

einem Teilbandsignal.

2. Vorrichtung nach Anspruch 1, bei der die Einrichtung zum Untersuchen (106a, 106b) ausgebildet ist, um eine Autokorrelationsfunktion für jedes der zumindest zwei Teilbandsignale zu berechnen. 5
3. Vorrichtung nach Anspruch 1 oder 2, bei der die Einrichtung zum Untersuchen (106a, 106b) folgende Merkmale aufweist: 10
 - eine Einrichtung zum Bilden einer Hüllkurve eines Teilbandsignals;
 - eine Einrichtung zum Glätten der Hüllkurve des Teilbandsignals, um eine geglättete Hüllkurve zu erhalten; 15
 - eine Einrichtung zum Differenzieren der geglätteten Hüllkurve, um eine differenzierte Hüllkurve zu erhalten; 20
 - eine Einrichtung zum Begrenzen der differenzierten Hüllkurve auf positive Werte, um eine begrenzte Hüllkurve zu erhalten; und 25
 - eine Einrichtung zum Bilden einer Autokorrelationsfunktion der begrenzten Hüllkurve, um die Rhythmus-Rohinformationen (108a, 108b) zu erhalten. 30
4. Vorrichtung nach Anspruch 2 oder 3, bei der die Einrichtung zum Bewerten (110a, 110b) der Qualität ausgebildet ist, um als Signifikanzmaß ein Verhältnis eines Maximum der Autokorrelationsfunktion zu einem Mittelwert der Autokorrelationsfunktion zu verwenden. 35
5. Vorrichtung nach Anspruch 2 oder 3, bei der die Einrichtung zum Bewerten (110a, 110b) der Qualität ausgebildet ist, um als Signifikanzmaß ein Verhältnis aus einem arithmetischen Mittelwert der Rhythmus-Rohinformationen zu einem geometrischen Mittelwert der Rhythmus-Rohinformationen zu verwenden. 40
6. Vorrichtung nach Anspruch 4 oder 5, bei der die Einrichtung zum Bewerten (110a, 110b) der Qualität ausgebildet ist, um die Autokorrelationsfunktion lediglich innerhalb eines Tempobereichs auszuwerten, der sich von einer minimalen Verzögerung, um ein maximales Tempo zu erhalten, zu einer maximalen Verzögerung, um ein minimales Tempo zu erhalten, erstreckt. 50
7. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung zum Ermitteln (114) folgende Merkmale aufweist: 55

eine Einrichtung (114a) zum Ableiten eines Gewichtungsfaktors für ein Teilband unter Verwendung des Signifikanzmaßes für das Teilband;

eine Einrichtung (114b) zum Gewichten der Rhythmus-Rohinformationen des Teilbands unter Verwendung des Gewichtungsfaktors für das Teilband, um gewichtete Rhythmus-Rohinformationen für das Teilband zu erhalten, und zum Zusammenfassen der gewichteten Rhythmus-Rohinformationen des Teilbands mit gewichteten oder ungewichteten Rhythmus-Rohinformationen des anderen Teilbands, um die Rhythmusinformationen des Audiosignals zu erhalten.

8. Vorrichtung nach Anspruch 7, bei der die Einrichtung (114a) zum Ableiten eines Gewichtungsfaktors angeordnet ist, um für jedes Teilbandsignal einen relativen Gewichtungsfaktor abzuleiten, wobei eine Summe aus den Gewichtungsfaktoren für alle Teilbandsignale 1 ergibt.
9. Vorrichtung nach Anspruch 8, bei der die Einrichtung (114a) zum Ableiten eines Gewichtungsfaktors angeordnet ist, um einen Gewichtungsfaktor als Verhältnis aus dem Signifikanzmaß eines Teilbandsignals zu der Summe der Signifikanzmaße aller Teilbandsignale abzuleiten.
10. Vorrichtung nach Anspruch 9, bei der die Einrichtung (106a, 106b) zum Untersuchen eines Teilbandsignals angeordnet ist, um ein Teilbandsignal zu untersuchen, dessen Länge größer als 10 Sekunden ist.
11. Verfahren zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen des Audiosignals, mit folgenden Schritten:
 - Zerlegen des Audiosignals in zumindest zwei Teilbandsignale (104a, 104b);
 - Untersuchen (106a, 106b) eines Teilbandsignals hinsichtlich einer Periodizität in dem Teilbandsignal, um Rhythmus-Rohinformationen (108a, 108b) für das Teilbandsignal zu erhalten;
 - Bewerten (110a, 110b) einer Qualität der Periodizität der Rhythmus-Rohinformationen (108a) des Teilbandsignals (104a), um ein Signifikanzmaß (112a) für das Teilbandsignal zu erhalten; und
 - Ermitteln der Rhythmusinformationen des Audiosignals unter Berücksichtigung des Signifi-

kanzmaßes (112a) des Teilbandsignals und der Rhythmus-Rohinformationen (108a, 108b) von zumindest einem Teilbandsignal.

Claims

1. Apparatus for analyzing an audio signal with regard to rhythm information of the audio signal, comprising:

means (102) for dividing the audio signal into at least two sub-band signals (104a, 104b);

means for examining (106a, 106b) a sub-band signal with regard to a periodicity in the sub-band signal, to obtain rhythm raw-information (108a, 108b) for the sub-band signal;

means for evaluating (110a, 110b) a quality of the periodicity of the rhythm raw-information (108a) of the sub-band signal (104a) to obtain a significance measure (112a) for the sub-band signal; and

means (114) for establishing rhythm information of the audio signal under consideration of the significance measure (112a) of the sub-band signal and the rhythm raw-information (108a, 108b) of at least one sub-band signal.

2. Apparatus according to claim 1, wherein the means for examining (106a, 106b) is formed to calculate an autocorrelation function for each of the least two sub-band signals.

3. Apparatus according to claim 1 or 2, wherein the means for examining (106a, 106b) comprises:

means for forming an envelope of a sub-band signal;

means for smoothing the envelope of the sub-band signal to obtain a smoothed envelope;

means for differentiating the smoothed envelope to obtain a differentiated envelope;

means for limiting the differentiated envelope to positive values to obtain a limited envelope; and

means for forming an autocorrelation function of the limited envelope to obtain the rhythm raw-information (108a, 108b).

4. Apparatus according to claim 2 or 3, wherein the means for evaluating (110a, 110b) of the quality is

formed to use a ratio of a maximum of the autocorrelation function to an average of the autocorrelation function as a significance measure.

5. Apparatus according to claim 2 or 3, wherein the means for evaluating (110a, 110b) of the quality is formed to use a ratio of an arithmetic average of the rhythm raw-information to a geometrical average of the rhythm raw-information as significance measure.

6. Apparatus according to claim 4 or 5, wherein the means for evaluating (110a, 110b) the quality is formed to evaluate the autocorrelation function merely within a tempo range, which extends from a minimum lag to obtain a maximum tempo to a maximum lag to obtain a minimum tempo.

7. Apparatus according to one of the previous claims, wherein means for establishing (114) comprises:

means (114a) for deriving a weighting factor for a sub-band by using the significance measure for the sub-band;

means (114b) for weighting a rhythm raw-information of the sub-band by using the weighting factor for the sub-band to obtain weighted rhythm raw-information for the sub-band and for summarizing the weighted rhythm raw-information of the sub-band with weighted or unweighted rhythm raw-information of the other sub-band to obtain the rhythm information of the audio signal.

8. Apparatus according to claim 7, wherein the means (114a) for deriving a weighting factor is disposed to derive a relative weighting factor for every sub-band signal, wherein a sum of the weighting factors for all sub-band signals equals 1.

9. Apparatus according to claim 8, wherein the means (114a) for deriving a weighting factor is disposed to derive a weighting factor as ratio of the significance measure of a sub-band signal to the sum of the significance measure of all sub-band signals.

10. Apparatus according to claim 9, wherein the means (106a, 106b) for examining a sub-band signal is disposed to examine a sub-band signal whose length is higher than 10 seconds.

11. Method for analyzing an audio signal with regard to rhythm information of the audio signal, comprising:

dividing the audio signal into at least two sub-band signals (104a, 104b);

examining (106a, 106b) a sub-band signal with

regard to a periodicity in the sub-band signal to obtain rhythm raw-information (108a, 108b) for the sub-band signal;

evaluating (110a, 110b) a quality of the periodicity of the rhythm raw-information (108a) of the sub-band signal (104a) to obtain a significance measure (112a) for the sub-band signal; and

establishing the rhythm information of the audio signal under consideration of the significance measure (112a) of the sub-band signal and the rhythm raw-information (108a, 108b) of at least one sub-band signal.

Revendications

1. Dispositif pour analyser un signal audio quant aux informations de rythme du signal audio, aux caractéristiques suivantes :
 - un dispositif (102) destiné à décomposer le signal audio en au moins deux signaux de bande partielle (104a, 104b) ;
 - un dispositif destiné à examiner (106a, 106b) un signal de bande partielle quant à une périodicité dans le signal de bande partielle, pour obtenir des informations de rythme brutes (108a, 108b) pour la bande partielle ;
 - un dispositif destiné à évaluer (110a, 110b) une qualité de la périodicité des informations de rythme brutes (108a) du signal de bande partielle (104a), pour obtenir une mesure de signification (112a) pour le signal de bande partielle ; et
 - un dispositif (114) destiné à déterminer les informations de rythme du signal audio en tenant compte de la mesure de signification (112a) du signal de bande partielle et des informations de rythme brutes (108a, 108b) d'au moins un signal de bande partielle.
2. Dispositif selon la revendication 1, dans lequel le dispositif destiné à examiner (106a, 106b) est réalisé de manière à calculer une fonction d'autocorrélation pour chacun des au moins deux signaux de bande partielle.
3. Dispositif selon la revendication 1 ou 2, dans lequel le dispositif destiné à examiner (106a, 106b) présente les caractéristiques suivantes :
 - un dispositif destiné à former une courbe enveloppante d'un signal de bande partielle ;
 - un dispositif destiné à aplanir la courbe enveloppante du signal de bande partielle, pour ob-

tenir une courbe enveloppante aplanie ;
 un dispositif destiné à différencier la courbe enveloppante aplanie, pour obtenir une courbe enveloppante différenciée ;
 un dispositif destiné à limiter la courbe enveloppante différenciée à des valeurs positives, pour obtenir une courbe enveloppante limitée, et
 un dispositif destiné à former une fonction d'autocorrélation de la courbe enveloppante, pour obtenir les informations de rythme brutes (108a, 108b).

4. Dispositif selon la revendication 2 ou 3, dans lequel le dispositif destiné à évaluer (110a, 110b) la qualité est réalisé de manière à utiliser, comme mesure de signification, un rapport entre un maximum de la fonction d'autocorrélation et une valeur moyenne de la fonction d'autocorrélation.
5. Dispositif selon la revendication 2 ou 3, dans lequel le dispositif destiné à évaluer (110a, 110b) la qualité est réalisé de manière à utiliser, comme mesure de signification, un rapport entre une valeur moyenne arithmétique des informations de rythme brutes et une valeur moyenne géométrique des informations de rythme brutes.
6. Dispositif selon la revendication 4 ou 5, dans lequel le dispositif destiné à évaluer (110a, 110b) la qualité est réalisé de manière à évaluer la fonction d'autocorrélation uniquement dans une plage de rythmes s'étendant d'un retard minimal pour obtenir un rythme maximal à un retard maximal pour obtenir un rythme minimal.
7. Dispositif selon l'une des revendications précédentes, dans lequel le dispositif destiné à déterminer (114) présente les caractéristiques suivantes :
 - un dispositif (114a) destiné à dériver un facteur de pondération pour une bande partielle à l'aide de la mesure de signification pour la bande partielle ;
 - un dispositif (114b) destiné à pondérer les informations de rythme brutes de la bande partielle à l'aide du facteur de pondération pour la bande partielle, pour obtenir les informations de rythme brutes pondérées et pour assembler les informations de rythme brutes pondérées de la bande partielle avec des informations de rythme brutes pondérées ou non pondérées de l'autre bande partielle, pour obtenir les informations de rythme du signal audio.
8. Dispositif selon la revendication 7, dans lequel le dispositif (114a) destiné à dériver un facteur de pondération est disposé de manière à dériver pour cha-

que signal de bande partielle un facteur de pondération relatif, une somme des facteurs de pondérations pour tous les signaux de bande partielle donnant 1.

5

9. Dispositif selon la revendication 8, dans lequel le dispositif (114a) destiné à dériver un facteur de pondération est disposé de manière à dériver un facteur de pondération comme rapport entre la mesure de signification d'un signal de bande partielle et la somme des mesures de signification de tous les signaux de bande partielle. 10
10. Dispositif selon la revendication 9, dans lequel le dispositif (106a, 106b) destiné à examiner un signal de bande partielle est disposé de manière à examiner un signal de bande partielle dont la longueur est supérieure à 10 secondes. 15
11. Procédé pour analyser un audio de signal quant aux informations de rythme du signal audio, aux étapes suivantes consistant à : 20

décomposer le signal audio en au moins deux signaux de bande partielle (104a, 104b) ; 25

examiner (106a, 106b) un signal de bande partielle quant à une périodicité dans le signal de bande partielle, pour obtenir des informations de rythme brutes (108a, 108b) pour le signal de bande partielle ; 30

évaluer (110a, 110b) une qualité de la périodicité des informations de rythme brutes (108a) du signal de bande partielle (104a), pour obtenir une mesure de signification (112a) du signal de bande partielle; et 35

déterminer les informations de rythme du signal audio compte tenu de la mesure de signification (112a) du signal de bande partielle et des informations de rythme brutes (108a, 108b) d'au moins un signal de bande partielle. 40

45

50

55

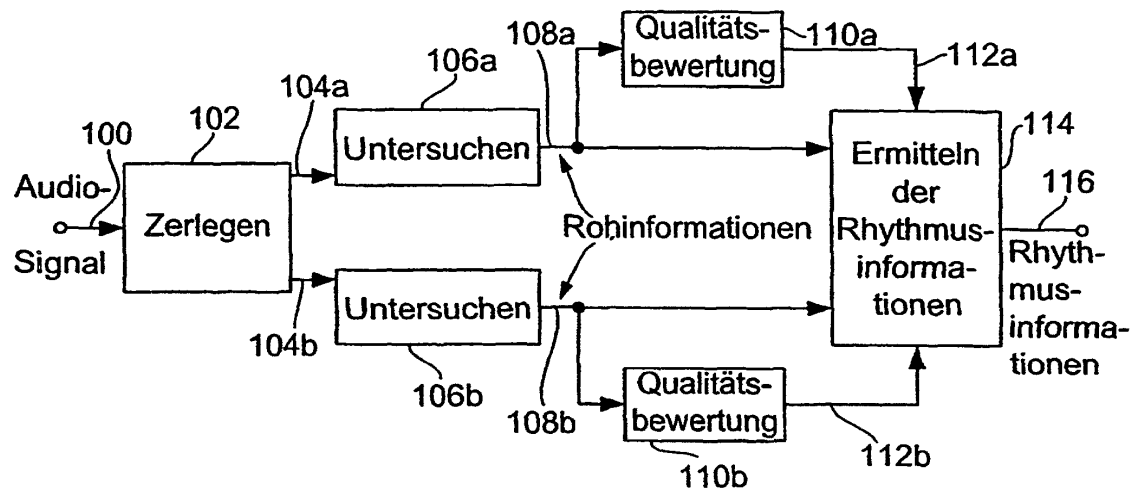


Fig. 1

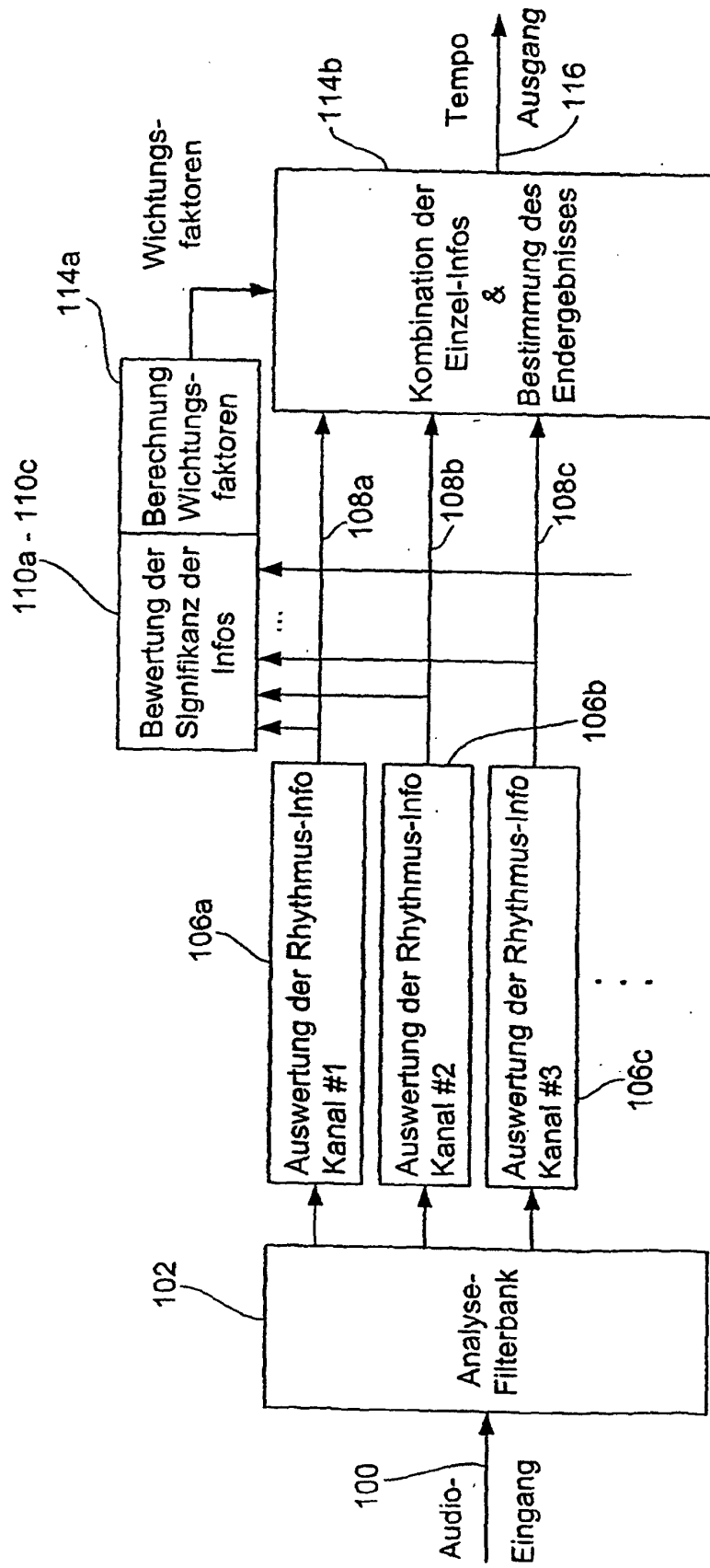


Fig. 2

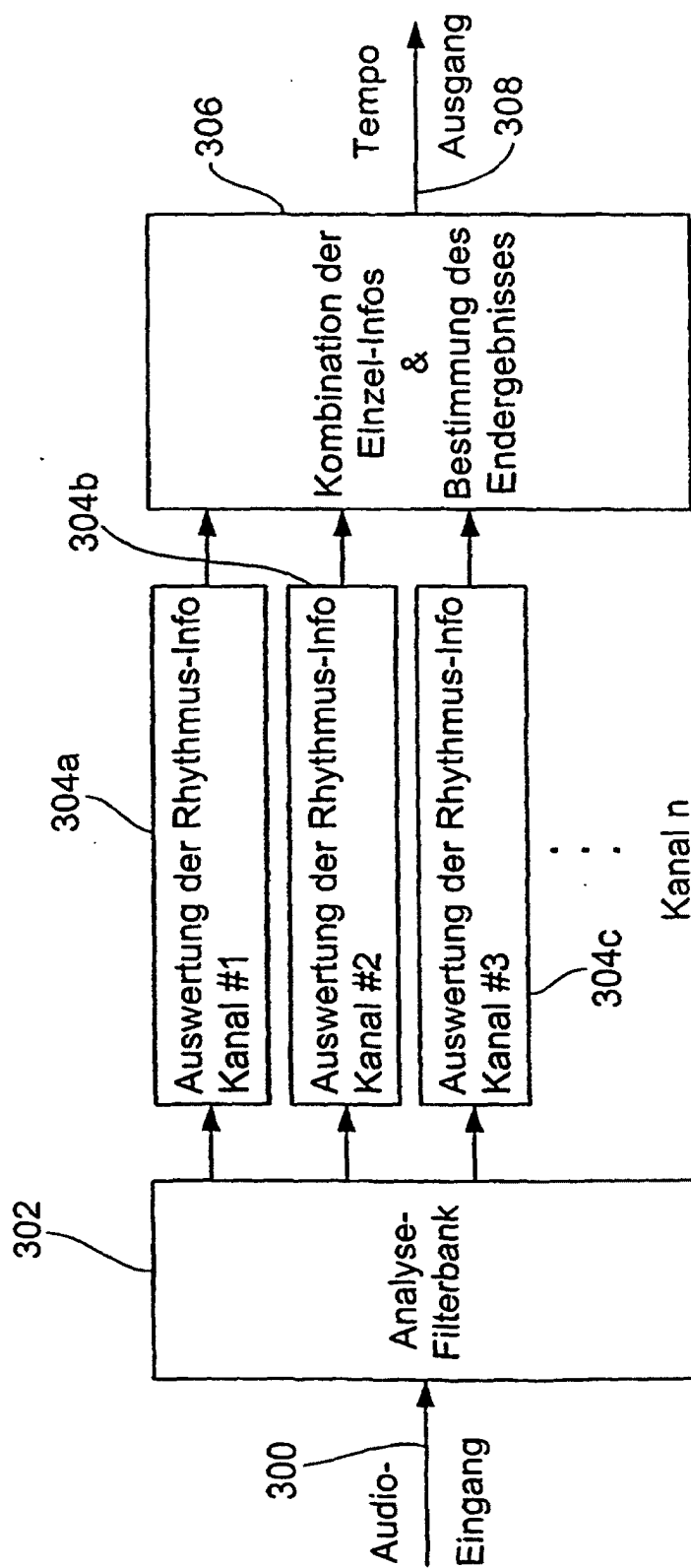


Fig. 3 (Stand der Technik)

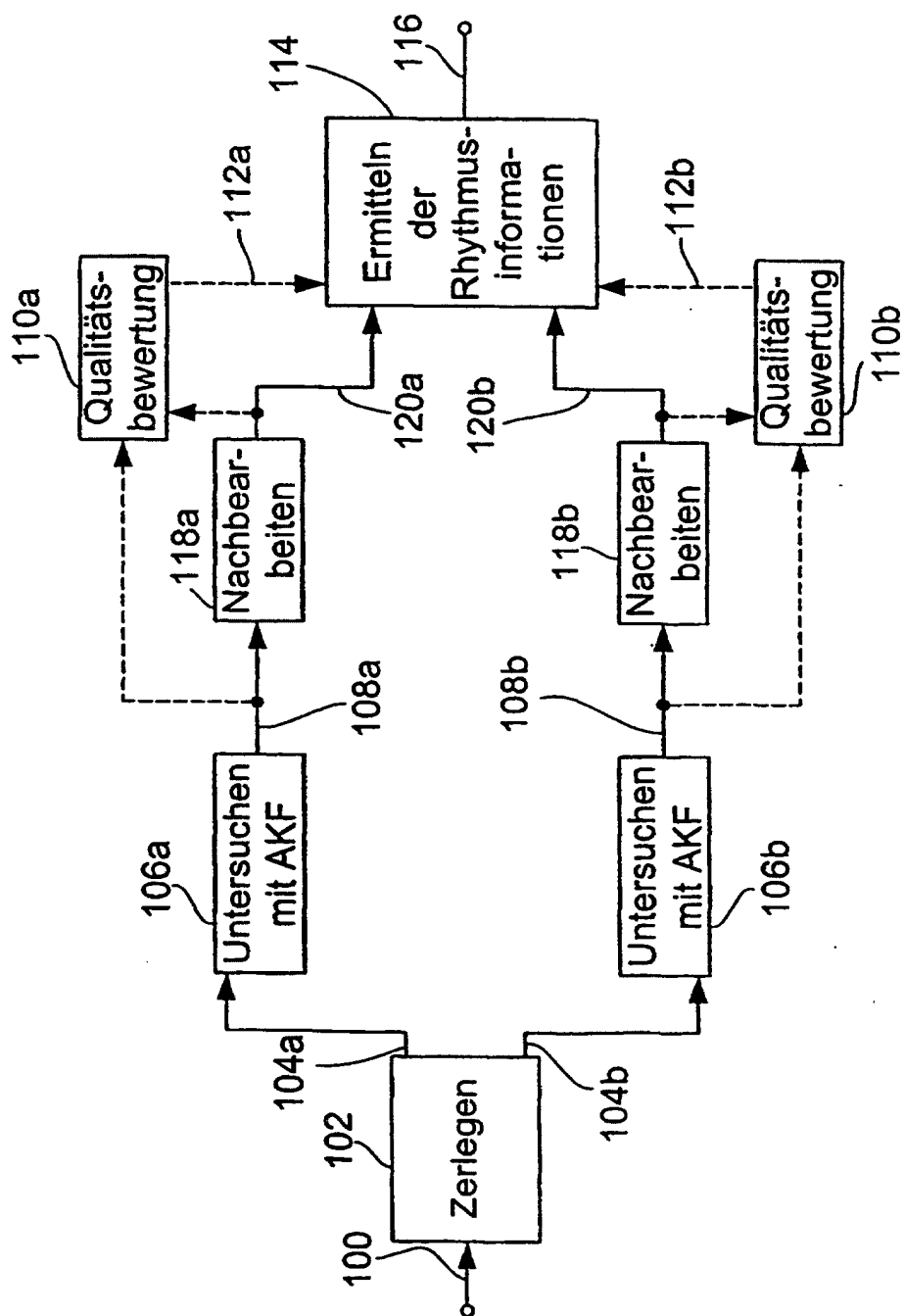


Fig. 4

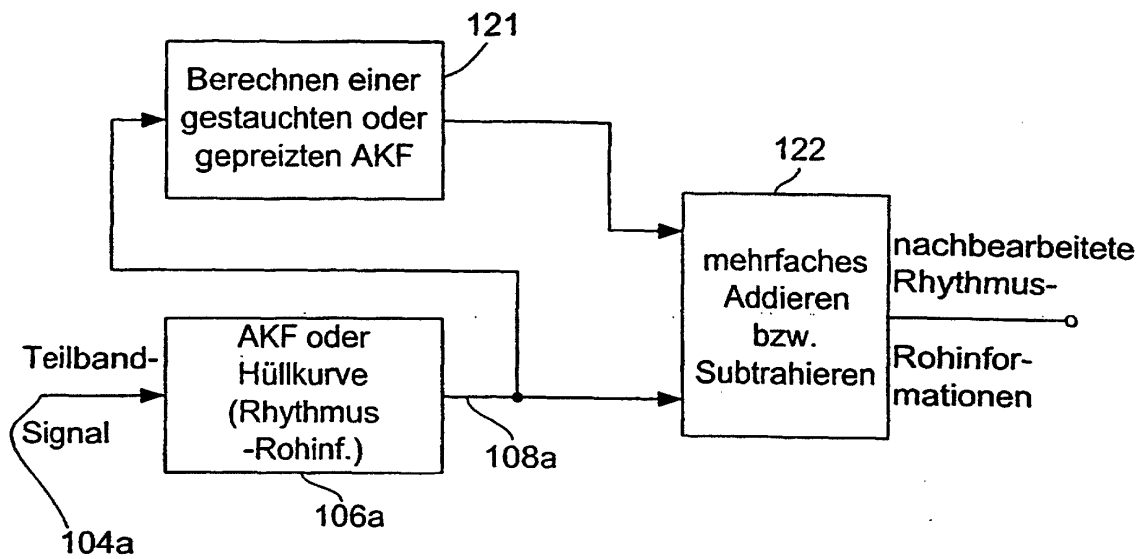


Fig. 5