(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 743 324 B1

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
31.10.2007 Patentblatt 2007/44

(21) Anmeldenummer: **05744658.5**

(22) Anmeldetag: **29.04.2005**

(51) Int Cl.:
G10L 11/00 *(2006.01)*

(86) Internationale Anmeldenummer:
PCT/EP2005/004685

(87) Internationale Veröffentlichungsnummer:
WO 2005/114651 (01.12.2005 Gazette 2005/48)

(54) **VORRICHTUNG UND VERFAHREN ZUM ANALYSIEREN EINES INFORMATIONSSIGNALS**

DEVICE AND METHOD FOR ANALYSING AN INFORMATION SIGNAL

DISPOSITIF ET PROCEDE POUR ANALYSER UN SIGNAL D'INFORMATION

(84) Benannte Vertragsstaaten:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU MC NL PL PT RO SE SI SK TR

(30) Priorität: **07.05.2004 DE 102004022660**

(43) Veröffentlichungstag der Anmeldung:
17.01.2007 Patentblatt 2007/03

(73) Patentinhaber: **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.**
80686 München (DE)

(72) Erfinder:
- **DITTMAR, Christian**
  98693 Ilmenau (DE)
- **UHLE, Christian**
  91052 Erlangen (DE)
- **HERRE, Jürgen**
  91054 Buckenhof (DE)

(74) Vertreter: **Zinkler, Franz et al**
Schoppe, Zimmermann, Stöckeler & Zinkler
Postfach 246
82043 Pullach bei München (DE)

(56) Entgegenhaltungen:
WO-A-01/16937

- **HEITTOLA T ET AL.: "Locating Segments with Drums in Music Signals" 3RD INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL ISMIR 2002, [Online] Oktober 2002 (2002-10), Seiten 1-6, XP002341222 Paris (Frankreich) Gefunden im Internet: URL:http: //www.cs.tut.fi/sgn/arg/music/ism ir2002long_ heittola.pdf> [gefunden am 2005-08-18]**
- **JARINA R ET AL: "Rhythm detection for speech-music discrimination in mpeg compressed domain" DIGITAL SIGNAL PROCESSING, 2002. DSP 2002. 2002 14TH INTERNATIONAL CONFERENCE ON SANTORINI, GREECE 1-3 JULY 2002, PISCATAWAY, NJ, USA,IEEE, US, Bd. 1, 1. Juli 2002 (2002-07-01), Seiten 129-132, XP010599702 ISBN: 0-7803-7503-3**

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 1 743 324 B1

Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf die Analyse von Informationssignalen, wie beispielsweise Audiosignalen und insbesondere auf die Analyse von Informationssignalen, die aus einer Überlagerung von Teilsignalen bestehen, wobei ein Teilsignal von einer Einzelquelle oder einer Gruppe von Einzelquellen stammen kann.

[0002] Die fortschreitende Entwicklung digitaler Distributionsmedien für multimediale Inhalte führt zu einer großen Vielfalt angebotener Daten. Für den menschlichen Nutzer ist die Grenze des Überschaubaren dabei längst überschritten. So gewinnt die inhaltliche Beschreibung der Daten durch Metadaten zunehmend an Bedeutung. Grundsätzlich besteht das Ziel, nicht nur Text-Dateien sondern auch z. B. Musikdateien, Videodateien oder sonstige Informationssignaldateien durchsuchbar zu machen, wobei derselbe Komfort wie bei gängigen Textdatenbanken angestrebt wird. Ein Ansatz hierfür ist der bekannte MPEG 7-Standard.

[0003] Insbesondere bei der Analyse von Audiosignalen, also Signalen, die Musik und/oder Sprache umfassen, ist die Extraktion von Fingerabdrücken von großer Bedeutung.

[0004] Angestrebt wird ferner, Audiodaten mit Metadaten "anzureichern", um für z. B. ein Musikstück auf der Basis eines Fingerabdrucks Metadaten wiederzugewinnen. Der "Fingerabdruck" soll einerseits aussagekräftig sein, und andererseits möglichst kurz und prägnant sein. "Fingerabdruck" bezeichnet somit ein aus einem Musik-Signal generiertes komprimiertes Informationssignal, welches nicht die Metadaten enthält, sondern zur Referenzierung auf die Metadaten z.B. durch Suche in einer Datenbank dient, z.B. in einem System zur Identifikation von Audiomaterial ("AudioID").

[0005] Normalerweise bestehen Musikdaten aus der Überlagerung von Teilsignalen von Einzelquellen. Während es bei einer Popmusik typischerweise relativ wenige Einzelquellen gibt, nämlich den Sänger, die Gitarre, die Bassgitarre, das Schlagzeug und ein Keyboard, so kann die Anzahl von Quellen für ein Orchesterstück sehr groß werden. Ein Orchesterstück und ein Popmusikstück beispielsweise bestehen aus einer Überlagerung der von den einzelnen Instrumenten abgegebenen Töne. Ein Orchesterstück bzw. irgendein Musikstück stellt somit eine Überlagerung von Teilsignalen von einzelnen Quellen dar, wobei die Teilsignale die von den einzelnen Instrumenten des Orchesters bzw. Popmusikensembles erzeugten Töne sind, und wobei die einzelnen Instrumente Einzelquellen sind.

[0006] Alternativ können auch Gruppen von ursprünglichen Quellen als Einzelquellen aufgefasst werden, so dass einem Signal wenigstens zwei Einzelquellen zugewiesen werden können.

[0007] Eine Analyse eines allgemeinen Informationssignals wird nachfolgend lediglich beispielhaft anhand eines Orchestersignals dargestellt. Die Analyse eines Orchestersignals kann auf vielerlei Arten und Weisen durchgeführt werden. So kann der Wunsch bestehen, die einzelnen Instrumente zu erkennen und aus dem Gesamtsignal die Einzelsignale der Instrumente zu extrahieren und gegebenenfalls in eine Notenschrift umzusetzen, wobei die Notenschrift als "Metadaten" fungieren würde. Weitere Möglichkeiten der Analyse bestehen darin, einen dominanten Rhythmus zu extrahieren, wobei eine Rhythmusextraktion auf der Basis der Schlaginstrumente besser vonstatten geht als auf der Basis der eher Ton-gebenden Instrumente, die auch als harmonisch-ausgehaltene oder "harmonic sustained" Instrumente bezeichnet werden. Während Schlaginstrumente typischerweise Pauken, Schlagzeuge, Rasseln oder sonstige Percussion-Instrumente umfassen, gehören zu den harmonisch ausgehaltenen Instrumenten alle sonstigen Instrumente, wie beispielsweise Geigen, Blasinstrumente, etc.

[0008] Weiterhin werden zu den Schlaginstrumenten alle jene akustischen oder synthetischen Klangerzeuger gezählt, die aufgrund ihrer Klangeigenschaften zur Rhythmussektion beitragen (z.B. Rhythmusgitarre).

[0009] So wäre es beispielsweise zur Rhythmusextraktion eines Musikstücks wünschenswert, aus dem gesamten Musikstück lediglich perkussive Anteile zu extrahieren und eine Rhythmuserkennung dann auf der Basis dieser percussiven Anteile durchzuführen, ohne dass die Rhythmuserkennung durch Signale von den harmonisch ausgehaltenen Instrumenten "gestört" wird.

[0010] Andererseits wird jegliche Analyse mit dem Ziel der Extraktion von Metadaten, die ausschließlich Informationen der harmonisch ausgehaltenen Instrumente benötigt (z.B. eine harmonische oder melodische Analyse), profitieren von einer vorgeschalteten Separation und einer Weiterverarbeitung der harmonisch ausgehaltenen Anteile.

[0011] In jüngster Vergangenheit wurde in diesem Zusammenhang über die Verwendung der Technik der Blind Source Separation (BSS) und der Independent Component Analysis (ICA) zur Signalverarbeitung und Signalanalyse berichtet. Einsatzgebiete finden sich insbesondere in der biomedizinischen Technik, der Kommunikationstechnik, der künstlichen Intelligenz und der Bildverarbeitung.

[0012] Allgemein umfasst der Begriff BSS Techniken zur Separation von Signalen aus einem Mix von Signalen mit einem Minimum an Vorkenntnissen über die Natur der Signale und des Mischungsprozesses. Die ICA ist ein Verfahren, welches sich der Annahme bedient, dass die einem Mix zugrundeliegenden Quellen zumindest bis zu einem gewissen Grad statistisch unabhängig voneinander sind. Weiterhin wird der Mischungsprozess als zeitlich unveränderlich und die Anzahl der beobachteten Mischsignale nicht kleiner als die Anzahl der der Mischung zugrundeliegenden Quellsignale angenommen.

[0013] Eine Erweiterung der ICA stellt die Independent Subspace Analysis (ISA) dar. Hier werden die Komponenten unterteilt in unabhängige Unterräume oder Sub-

spaces, deren Komponenten nicht statistisch unabhängig sein müssen. Durch eine Transformation des Musiksignals wird eine mehrdimensionale Darstellung des Mischsignals ermittelt und der letzten Annahme für die ICA entsprochen. Verschiedene Verfahren zur Berechnung der unabhängigen Komponenten wurden in den letzten Jahren entwickelt. Einschlägige Literaturstellen, die sich zum Teil auch mit der Analyse von Audiosignalen beschäftigen, sind folgende:

[1] M.A. Casey and A. Westner, "Separation of Mixed Audio Sources by Independent Subspace Analysis", in Proc. of the International Computer Music Conference, Berlin, 2000

[2] I.F.O. Orife, "Riddim: A rhythm analysis and decomposition tool based on independent subspace analysis", Master thesis, Darthmouth College, Hanover, New Hampshire, 2001

[3] C. Uhle, C. Dittmar and T. Sporer, "Extraction of Drum Tracks from polyphonic Music using Independent Subspace Analysis", in Proc. of the Fourth International Symposium on Independent Component Analysis, Nara, Japan, 2003

[4] D. Fitzgerald, B. Lawlor and E. Coyle, "Prior Subspace Analysis for Drum Transcription", in Proc. Of the 114th AES Convention, Amsterdam, 2003

[5] D. Fitzgerald, B. Lawlor and E. Coyle,"Drum Transcription in the presence of pitched instruments using Prior Subspace Analysis", in Proc. of the ISSC, Limerick, Ireland, 2003

[6] M. Plumbley, "Algorithms for Non-Negative Independent Component Analysis", in IEEE Transactions on Neural Networks, 14 (3), pp 534- 543, May 2003

[7] T. HEITTOLA ET AL.: "Locating Segments with Drums in Music Signals" 3RD INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL ISMIR 2002, Oktober 2002 (2002-10), Seiten 1-6, XP002341222

[0014] In [1] wird ein Verfahren zur Separation von Einzelquellen aus Mono-Audiosignalen dargestellt. In [2] wird eine Anwendung für eine Auftrennung in Einzelspuren und anschließend die Rhythmusanalyse gegeben. In [3] wird eine Komponentenanalyse durchgeführt, um eine Auftrennung in perkussive und nicht-perkussive Klänge eines polyphonen Stücks zu erreichen. In [4] wird die Independent Component Analysis (ICA) auf Amplitudenbasen angewendet, die mittels allgemein berechneter Frequenzbasen aus einer Spektrogrammdarstellung einer Schlagzeugspur gewonnen werden. Dies wird zum Zwecke der Transkription durchgeführt. In [5] wird dieses Verfahren auf polyphone Musikstücke erweitert.

[0015] Die erste vorstehend genannte Veröffentlichung von Casey sei nachfolgend beispielhaft für den Stand der Technik dargestellt. Diese Veröffentlichung beschreibt ein Verfahren zum Trennen von gemischten Audioquellen durch die Technik der unabhängigen Unterraumanalyse. Hierzu wird unter Verwendung von BSS-Techniken ein Audiosignal in einzelne Komponentensignale aufgesplittet. Zum Feststellen, welche der einzelnen Komponentensignale zu einem Multikomponenten-Unterraum gehören, wird eine Gruppierung dahingehend durchgeführt, dass die Ähnlichkeit der Komponenten untereinander durch ein sogenanntes Ixegramm dargestellt wird. Das Ixegramm wird als Kreuzentropie-Matrix der unabhängigen Komponenten untereinander bezeichnet. Es wird dadurch berechnet, dass alle einzelnen Komponentensignale paarweise in einer Korrelationsberechnung untersucht werden, um ein Maß dafür zu finden, wie ähnlich zwei Komponenten sind. Über allen Komponentensignalen wird daher eine erschöpfende paarweise Ähnlichkeitsberechnung durchgeführt, so dass sich eine Ähnlichkeitsmatrix ergibt, bei der entlang einer y-Achse alle Komponentensignale aufgetragen werden, und bei der ferner entlang der x-Achse ebenfalls alle Komponentensignale aufgetragen werden. Dieses zweidimensionale Array liefert für jedes Komponentensignal ein Ähnlichkeitsmaß mit jeweils einem anderen Komponentensignal. Das Ixegramm, also die zweidimensionale Matrix, wird nun dazu verwendet, ein Clustering durchzuführen, wobei hierzu eine Gruppierung unter Verwendung eines Cluster-Algorithmus auf der Basis von diadischen Daten durchgeführt wird. Um eine optimale Partitionierung des Ixegramms in k Klassen durchzuführen, wird eine Kostenfunktion definiert, die die Kompaktheit innerhalb eines Clusters misst und die Homogenität zwischen Clustern bestimmt. Die Kostenfunktion wird minimiert, so dass sich letztendlich eine Zuordnung von einzelnen Komponenten zu einzelnen Unterräumen ergibt. Angewendet auf ein Signal, das einen Sprecher im Kontext eines durchgehenden Wasserfallrauschens darstellt, ergibt sich als Unterraum der Sprecher, wobei das rekonstruierte Informationssignal des Sprecher-Unterraums eine signifikante Dämpfung des Wasserfallrauschens zeigt.

[0016] Nachteilig an den beschriebenen Konzepten ist die Tatsache, dass sehr wahrscheinlich der Fall auftritt, dass die Signalanteile einer Quelle auf verschiedenen Komponentensignalen zu liegen kommen. Dies ist der Grund dafür, dass, wie es vorstehend ausgeführt worden ist, eine komplexe und rechenzeitintensive Ähnlichkeitsberechnung unter allen Komponentensignalen durchgeführt wird, um die zweidimensionale Ähnlichkeitsmatrix zu erhalten, auf der Basis derer dann mittels einer zu minimierenden Kostenfunktion letztendlich eine Einteilung von Komponentensignalen in Unterräume durchgeführt wird.

[0017] Des weiteren nachteilhaft ist, dass in dem Fall, in dem mehrere einzelne Quelle vorliegen, also wo das

Ausgangssignal nicht a priori bekannt ist, zwar nach längerer Rechnung eine Ähnlichkeitsverteilung vorliegt, dass jedoch die Ähnlichkeitsverteilung selbst noch keinen tatsächlichen Einblick in die tatsächliche Audioszene liefert. So weiß der Betrachter lediglich, dass bestimmte Komponentensignale zueinander im Hinblick auf die minimierte Kostenfunktion ähnlich sind. Er weiß jedoch nicht, welche Informationen diese letztendlich erhaltenen Unterräume tragen bzw. welche ursprüngliche Einzelquelle oder welche Gruppe von Einzelquellen durch einen Unterraum dargestellt werden.

[0018] Die Independent Subspace Analysis (ISA) kann somit dazu genutzt werden, um eine Zeit-Frequenz-Repräsentation, z.B. ein Spektrogramm, eines Audiosignals in unabhängige Komponenten-Spektren zu zerlegen. Die vorher beschriebenen bisherigen Verfahren setzen dabei entweder auf eine berechnungsintensive Ermittlung von Frequenz- und Amplitudenbasen aus dem gesamten Spektrogramm oder auf a priori definierte Frequenzbasen. Solche a priori definierte Frequenzbasen bzw. Profilspektren bestehen beispielsweise darin, dass man sagt, dass in einem Stück sehr wahrscheinlich eine Trompete vorkommt, und dass dann zur Signalanalyse ein Beispiel-Spektrum einer Trompete verwendet wird.

[0019] Dieses Prozedere hat den Nachteil, dass man von vornherein alle auftretenden Instrumente wissen muss, was der automatisierten Verarbeitung bereits prinzipiell zuwiderläuft. Ein weiterer Nachteil besteht darin, dass es, wenn genau gearbeitet werden soll, nicht nur zum Beispiel Trompeten gibt, sondern viele verschiedene Arten von Trompeten, die sich alle in ihrer Klangfarbe und damit in ihrem Spektrum unterscheiden. Würde man derart vorgehen, dass man nun alle Arten von Beispielspektren zur Komponentenanalyse einsetzt, so wird das Verfahren wiederum sehr aufwändig und bekommt eine sehr hohe Redundanz, da typischerweise nicht alle denkbaren verschiedenen Trompeten in einem Stück vorkommen, sondern nur Trompeten einer einzigen Art, also mit einem einzigen Profilspektrum, oder vielleicht mit einigen wenigen unterschiedlichen Klangfarben, also wenigen Profilspektren. Noch problematischer wird es bei verschiedenen Noten einer Trompete, zumal jeder Ton je nach Tonhöhe ein aufgespreiztes/gestauchtes Profilspektrum aufweist. Um dies zu berücksichtigen ist der Rechenaufwand ebenfalls immens.

[0020] Andererseits ist die Zerlegung aufgrund von ISA-Konzepten dann außerordentlich rechenaufwändig und störungsanfällig, wenn das gesamte Spektrogramm verwendet wird. Es sei darauf hingewiesen, dass ein Spektrogramm typischerweise aus einer Folge von Einzelspektren besteht, wobei zwischen den Einzelspektren eine Hopping-Zeitdauer definiert ist, und wobei ein Spektrum eine bestimmte Anzahl von Abtastwerten repräsentiert, so dass einem Spektrum eine bestimmte zeitliche Länge, also ein Block von Abtastwerten des Signals zugeordnet ist. Typischerweise wird die Dauer, die durch den Block von Abtastwerten, aus dem ein Spektrum berechnet wird, repräsentiert wird, wesentlich größer als die Hopping-Zeit sein, um im Hinblick auf die erforderliche Frequenzauflösung und im Hinblick auf die erforderliche Zeitauflösung ein zufriedenstellendes Spektrogramm zu erhalten. Anderseits ist jedoch zu sehen, dass diese Spektrogrammdarstellung außerordentlich redundant ist. Wird beispielsweise der Fall betrachtet, dass eine Hopping-Zeitdauer 10 ms beträgt, und dass einem Spektrum ein Block von Abtastwerten mit einer zeitlichen Länge von zum Beispiel 100 ms zugrunde liegt, so tritt jeder Abtastwert in 10 aufeinander folgenden Spektren auf. Die dadurch erzeugte Redundanz kann insbesondere dann, wenn eine größere Anzahl von Instrumenten gesucht wird, die Rechenzeitanforderungen in astronomische Höhen treiben.

[0021] Des weiteren ist der Ansatz, auf der Basis des gesamten Spektrogramms zu arbeiten, für solche Fälle nachteilhaft, bei denen von einem Signal nicht alle enthaltenen Quellen extrahiert werden sollen, sondern nur zum Beispiel Quellen eines bestimmten Typs, also Quellen, die eine bestimmte Charakteristik haben. Eine solche Charakteristik kann perkussive Quellen, also Schlaginstrumente, betreffen, oder sogenannte gepitchte Instrumente, die auch als Harmonic-Sustained-Instrumente bezeichnet werden, welche typische Melodie-Instrumente, wie Trompete, Geige, etc. sind. Ein Verfahren, das auf der Basis aller dieser Quellen arbeitet, ist dann zu aufwändig und letztendlich auch zu wenig robust, wenn zum Beispiel nur einige Quellen, nämlich die Quellen, die eine bestimmte Charakteristik erfüllen sollen, extrahiert werden sollen. In diesem Fall werden nämlich Einzelspektren des Spektrogramms, in denen solche Quellen nicht oder nur sehr gering auftreten, das Gesamtergebnis verfälschen bzw. "verwaschen", da diese Spektren des Spektrogramms selbstverständlich ebenfalls in die letztendliche Komponentenanalyseberechnung eingehen wie die signifikanten Spektren.

[0022] In [7] wird ein weiteres Verfahren offenbart, um aus einem Musiksignal Abschnitte mit Tonsignalen eines Schlagzeuginstrumentes zu bestimmen.

[0023] Die Aufgabe der vorliegenden Erfindung besteht darin, ein robustes und rechenzeiteffizientes Konzept zum Analysieren eines Informationssignals zu schaffen.

[0024] Diese Aufgabe wird durch eine Vorrichtung zum Analysieren eines Informationssignals nach Patentanspruch 1, ein Verfahren zum Analysieren eines Informationssignals nach Patentanspruch 24 oder ein Computerprogramm nach Patentanspruch 25 gelöst.

[0025] Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, dass eine robuste und effiziente Informationssignalanalyse dadurch erreicht wird, dass zunächst signifikante Kurzzeitspektren oder von signifikanten Kurzzeitspektren abgeleitete Kurzzeitspektren, wie beispielsweise Differenzspektren etc. aus dem gesamten Informationssignal bzw. aus dem Spektrogramm des Informationssignals extrahiert werden, wobei solche Kurzzeitspektren extrahiert werden, die einer spezifischen Charakteristik näher kommen als andere Kurzzeitspek-

tren des Informationssignals.

[0026] Vorzugsweise werden Kurzzeitspektren extrahiert, die perkussive Anteile haben, und werden somit Kurzzeitspektren nicht extrahiert, die harmonische Anteile haben. In diesem Fall ist die spezifische Charakteristik eine perkussive bzw. Schlagzeug-Charakteristik.

[0027] Die extrahierten Kurzzeitspektren oder von den extrahierten Kurzzeitspektren abgeleitete Kurzzeitspektren werden dann eine Einrichtung zum Zerlegen der Kurzzeitspektren in Komponentensignalspektren zugeführt, wobei ein Komponentensignalspektrum ein Profilspektrum einer Tonquelle darstellt, die einen Ton erzeugt, der der gesuchten Charakteristik entspricht, und wobei ein anderes Komponentensignalspektrum ein anderes Profilspektrum einer Tonquelle darstellt, die einen Ton erzeugt, der ebenfalls der gesuchten Charakteristik entspricht.

[0028] Schließlich wird auf der Basis der Profilspektren der Tonquellen eine Amplitudenhüllkurve über der Zeit berechnet, wobei für die Berechnung der Amplitudenhüllkurve über der Zeit die ermittelten Profilspektren sowie die ursprünglichen Kurzzeitspektren verwendet werden, so dass für jeden Zeitpunkt, zu dem ein Kurzzeitspektrum abgenommen wurde, auch ein Amplitudenwert erhalten wird.

[0029] Die derart erhaltene Information, nämlich verschiedene Profil-Spektren sowie Amplitudenhüllkurven für die Profilspektren, liefern liefert eine vollständige Beschreibung des Musik- bzw. Informationssignals im Hinblick auf die spezifizierte Charakteristik, nach der extrahiert worden ist, so dass diese Informationen bereits ausreichend sein können, um eine Transkription zu machen, also um mit Konzepten der Merkmalsextraktion und Segmentierung zunächst festzustellen, welches Instrument zu dem Profilspektrum "gehört", und welche Rhythmik vorliegt, also welche Anstiege und Abfallereignisse vorliegen, die auf zu bestimmten Zeitpunkten gespielte Noten dieses Instruments hinweisen.

[0030] Die vorliegende Erfindung ist dahingehend vorteilhaft, dass zur Berechnung der Komponentenanalyse also zum Zerlegen nicht das gesamte Spektrogramm verwendet wird, sondern nur extrahierte Kurzzeitspektren, dass also die Berechnung der Independent Subspace Analysis (ISA) nur anhand einer Teilmenge aller Spektren stattfindet, so dass die Rechenanforderungen gesenkt werden. Ferner wird auch die Robustheit hinsichtlich des Auffindens bestimmter Quellen erhöht, zumal andere Kurzzeitspektren, die die spezifizierte Charakteristik nicht erfüllen, bei der Komponentenanalyse nicht vorhanden sind und damit auch keine Störung bzw. keine "Verwaschung" der tatsächlichen Spektren darstellen.

[0031] Darüber hinaus ist das erfindungsgemäße Konzept dahingehend vorteilhaft, dass die Profilspektren direkt aus dem Signal ermittelt werden, ohne dass sich die Problematik der vorgefertigten Profilspektren ergibt, welche wiederum zu entweder ungenauen Ergebnissen oder zu einem erhöhten Rechenaufwand führen würde.

[0032] Vorzugsweise wird zur Detektion und Klassifikation von perkussiven, nicht-harmonischen Instrumenten in polyphonen Audiosignalen das erfindungsgemäße Konzept eingesetzt, um sowohl Profilspektren als auch Amplitudenhüllkurven für die einzelnen Profilspektren zu erhalten.

[0033] Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend bezugnehmend auf die beiliegenden Zeichnungen detailliert erläutert. Es zeigen:

Fig. 1 ein Blockschaltbild der erfindungsgemäßen Vorrichtung zum Analysieren eines Informationssignals;

Fig. 2 ein Blockschaltbild einer bevorzugten Ausführungsform der erfindungsgemäßen Vorrichtung zum Analysieren eines Informationssignals;

Fig. 3a ein Beispiel für eine Amplitudenhüllkurve für eine perkussive Quelle;

Fig. 3b ein Beispiel für ein Profilspektrum für eine perkussive Quelle;

Fig. 4a ein Beispiel für eine Amplitudenhüllkurve für ein harmonisch ausgehaltenes Instrument; und

Fig. 4b ein Beispiel für ein Profilspektrum für ein harmonisch ausgehaltenes Instrument.

[0034] Fig. 1 zeigt ein bevorzugtes Ausführungsbeispiel für eine erfindungsgemäße Vorrichtung zum Analysieren eines Informationssignals, das über eine Eingangsleitung 10 einer Einrichtung 12 zum Bereitstellen einer Folge von Kurzzeitspektren, die das Informationssignal darstellen, zugeführt wird. Wie es durch eine Umwegleitung 14 in Fig. 1 dargestellt ist, die gestrichelt gezeichnet ist, kann das Informationssignal auch zum Beispiel in zeitlicher Form einer Einrichtung 16 zum Extrahieren von signifikanten Kurzzeitspektren oder von den Kurzzeitspektren abgeleiteten Kurzzeitspektren aus dem Informationssignal zugeführt werden, wobei die Einrichtung zum Extrahieren ausgebildet ist, um solche Kurzzeitspektren zu extrahieren, die einer spezifischen Charakteristik näher kommen als andere Kurzzeitspektren des Informationssignals.

[0035] Die extrahierten Spektren, also die ursprünglichen Kurzzeitspektren oder die von den ursprünglichen Kurzzeitspektren zum Beispiel durch Differenzieren, Differenzieren und Gleichrichten oder durch andere Operationen abgeleiteten Kurzzeitspektren werden einer Einrichtung 18 zum Zerlegen der extrahierten Kurzzeitspektren in KomponentenSignalspektren zugeführt, wobei ein KomponentenSignalspektrum ein Profilspektrum einer Tonquelle darstellt, die einen Ton erzeugt, der

der gesuchten Charakteristik entspricht, und wobei ein anderes Profilspektrum eine andere Tonquelle darstellt, die einen Ton erzeugt, der der gesuchten Charakteristik ebenfalls entspricht.

[0036] Die Profilspektren werden schließlich einer Einrichtung 20 zum Berechnen einer Amplitudenhüllkurve für die eine Tonquelle zugeführt, wobei die Amplitudenhüllkurve angibt, wie sich die Profilspektren einer Tonquelle über der Zeit ändern, und insbesondere wie sich die Intensität oder Gewichtung eines Profilspektrums mit der Zeit ändert. Die Einrichtung 20 ist ausgebildet, um auf der Basis der Folge von Kurzzeitspektren einerseits sowie auf der Basis der Profilspektren andererseits zu arbeiten, wie es aus Fig. 1 hervorgeht. Ausgangsseitig liefert die Einrichtung 20 zum Berechnen Amplitudenhüllkurven für die Quellen, während die Einrichtung 18 Profilspektren für die Tonquellen liefert. Die Profilspektren sowie die zugehörigen Amplitudenhüllkurven liefern eine vollständige Beschreibung des Anteils des Informationssignals, der der spezifischen Charakteristik entspricht. Vorzugsweise ist dieser Anteil der perkussive Anteil eines Musikstücks. Alternativ könnte dieser Anteil jedoch auch der harmonische Anteil sein. In diesem Fall würde die Einrichtung zum Extrahieren von signifikanten Kurzzeitspektren anders ausgestaltet sein wie in dem Fall, in dem die spezifische Charakteristik eine perkussive Charakteristik ist.

[0037] Nachfolgend wird bezugnehmend auf Fig. 2 ein bevorzugtes Ausführungsbeispiel der vorliegenden Erfindung dargestellt. Vorzugsweise wird mit den Profilspektren F und den Amplitudenhüllkurven E eine Detektion und Klassifikation perkussiver, nicht-harmonischer Instrumente durchgeführt, wie es auch durch einen Block 22 in Fig. 2 dargestellt ist. Hierauf wird jedoch noch später eingegangen.

[0038] Wie es aus Fig. 2 ersichtlich ist, ist die Einrichtung 12 zum Bereitstellen einer Folge von Kurzzeitspektren ausgebildet, um mittels einer geeigneten Zeit-FrequenzTransformation ein Amplitudenspektrogramm **X** zu erzeugen. Die Zeit/Frequenz-Einrichtung 12 ist vorzugsweise eine Einrichtung zum Durchführen einer Kurzzeit-Fourier-Transformation mit einer bestimmten Hopping-Periode, oder umfasst Filterbanken. Optional wird auch ein Phasenspektrogramm als zusätzliche Informationsquelle gewonnen, wie es in Fig. 2 durch einen Phasenpfeil 13 dargestellt ist. Hierauf wird durch Differenzierung entlang der zeitlichen Ausdehnung jeder einzelnen Spektrogrammzeile, also jedes einzelnen Frequenz-Bins, ein Differenzspektrogramm $\dot{X}$ gewonnen, wie es durch den Differenzierer 16a dargestellt ist. Die durch Differenzierung entstandenen negativen Anteile werden auf Null gesetzt oder - alternativ - positiv gemacht. Damit ergibt sich ein nichtnegatives Differenzspektrogramm $\dot{X}$. Dieses nicht-negative

[0039] Differenzenspektrogramm wird einem Maximumsucher 16c zugeführt, der ausgebildet ist, um nach den Zeitpunkten t, also nach den Indizes der entsprechenden Spektrogrammspalten, des Auftretens lokaler Maxima in einer Detektionsfunktion e, die vor dem Maximumsucher 16c berechnet wird, zu suchen. Wie es später noch ausgeführt wird, kann die Detektions-funktion beispielsweise durch Aufsummierung über alle Zeilen von $\hat{X}$ und anschließende Glättung gewonnen werden.

[0040] Optional wird es bevorzugt, die Phaseninformationen, die über die Phasenleitung 13 vom Block 12 zum Block 16c geliefert werden, als Indikator für die Verlässlichkeit der gefundenen Maxima zu verwenden. Die Spektren, für die der Maximumsucher ein Maximum in der Detektionsfunktion erfasst, werden als $\hat{X}_t$ verwendet und stellen die extrahierten

[0041] Kurzzeitspektren dar.

[0042] In Block 18a wird eine Principle Component Analysis (PCA) durchgeführt. Hierbei wird zunächst eine gesuchte Anzahl von Komponenten d festgelegt. Dann wird die PCA nach einem geeigneten Verfahren wie beispielsweise Singular Value Decomposition oder Eigenvalue Decomposition über die Spalten der Matrix $\hat{X}_t$ durchgeführt.

$$\tilde{\mathbf{X}} = \hat{\mathbf{X}}_t \cdot \mathbf{T}$$

[0043] Die Transformationsmatrix T bewirkt eine Dimensionsreduktion auf $\tilde{X}$, was sich in einer Verringerung der Anzahl von Spalten dieser Matrix auswirkt. Ferner wird eine Dekorrelation und Varianznormierung erreicht. In Block 18b wird dann eine nicht-negative Independent Component Analysis ausgeführt. Hierbei wird das in [6] gezeigte Verfahren der nicht-negativen Independent Component Analysis auf $\tilde{X}$ zur Berechnung einer Separationsmatrix A ausgeführt. Gemäß der nachfolgenden Gleichung wird $\tilde{X}$ in unabhängige Komponenten zerlegt.

$$\mathbf{F} = \mathbf{A} \cdot \tilde{\mathbf{X}}$$

[0044] Unabhängige Komponenten F werden als statische spektrale Profile bzw. Profilspektren der auftretenden Klangquellen interpretiert. In einem Block 20 wird dann die Amplitudenbasis bzw. die Amplitudenhüllkurve E gemäß folgender Gleichung für die einzelnen Tonquellen extrahiert.

$$\mathbf{E} = \mathbf{F} \cdot \mathbf{X}$$

[0045] Die Amplitudenbasis wird als Satz von zeitveränderlichen Amplitudenhüllkurven der korrespondierenden spektralen Profile interpretiert.

[0046] Erfindungsgemäß wird das spektrale Profil aus

dem Musiksignal selbst gewonnen. Hierdurch wird die Rechenkomplexität gegenüber den bisherigen Verfahren reduziert, und es wird eine höhere Robustheit gegenüber stationären Signalanteilen, also Signalanteilen aufgrund von Harmonic Sustained Instrumenten erreicht.

[0047] In einem Block 22 wird dann eine Merkmalsextraktion und eine Klassifizierungsoperation durchgeführt. Insbesondere werden die Komponenten in zwei Teilmengen unterschieden, nämlich zunächst in eine Teilmenge mit den Eigenschaften nicht perkussiv, also quasi harmonisch, und in eine andere perkussive Teilmenge.. Darüber hinaus werden die Komponenten mit der Eigenschaft perkussiv/dissonant weiter in verschiedenen Instrumentenklassen klassifiziert.

[0048] Zur Einteilung in die zwei Teilmengen werden die Merkmale der Perkussivität bzw. spektralen Dissonanz verwendet.

[0049] Zur Instrumentenklassifikation werden folgende Merkmale eingesetzt:

> geglättete Version der spektralen Profile als Suchmuster in einer Trainingsdatenbank mit Profilen einzelner Instrumente, spektraler Zentroid, spektrale Ausbreitung, spektrale Schiefheit, Mittenfrequenzen, Intensitäten, Ausdehnung, Schiefheit der deutlichsten Partiallinien, ...

[0050] In folgende Instrumentenklassen kann beispielsweise klassifiziert werden:

> Kick Drum, Snare Drum, Hi-Hat, Cymbal, Tom, Bongo, Conga, Woodblock, Cowbell, Timbales, Shaker, Tabla, Tambourine, Triangle, Daburka, Castagnets, Handclaps.

[0051] In einem Block 24 kann dann zur weiteren Erhöhung des Robustheit des erfindungsgemäßen Konzepts ein Entscheidung für Schlagzeugeinsätze bzw. eine Annahme oder Akzeptanz von perkussiven Maxima durchgeführt werden. So werden Maxima mit einem transienten Anstieg in der Amplitudenhüllkurve über einem variablen Schwellwert als perkussives Ereignis angenommen, während Maxima mit einem transienten Anstieg unterhalb des variablen Schwellwerts verworfen werden bzw. als Artefakt erkannt und ignoriert werden. Der variable Schwellwert variiert vorzugsweise mit der Gesamtamplitude in einem größeren Bereich um das Maximum herum. Die Ausgabe erfolgt in einer geeigneten Form, die dem Zeitpunkt perkussiver Ereignisse eine Instrumentenklasse, eine Intensität und eventuell weitere Informationen zuordnet, wie beispielsweise Noten- bzw. Rhythmusinformationen im MIDI-Format.

[0052] An dieser Stelle sei darauf hingewiesen, dass die Einrichtung 16 zum Extrahieren von signifikanten Kurzzeitspektren ausgebildet sein kann, um diese Extraktion anhand tatsächlicher Kurzzeitspektren durchzuführen, wie sie zum Beispiel bei einer Kurzzeit-Fourier-Transformation erhalten werden. Insbesondere bei dem Anwendungsbeispiel der vorliegenden Erfindung, bei dem die spezifische Charakteristik die Schlagzeug-Charakteristik bzw. die perkussive Charakteristik ist, wird es bevorzugt, nicht tatsächliche Kurzzeitspektren zu extrahieren, sondern Kurzzeitspektren aus einem differenzierten Spektrogramm, also aus Differenzspektren. Die Differenzierung, wie sie in Block 16a in Fig. 2 gezeigt ist, führt die Folge von Kurzzeitspektren zu einer Folge von abgeleiteten bzw. differenzierten Spektren, wobei jedes (differenzierte) Kurzzeitspektrum nunmehr die Änderungen zwischen einem ursprünglichen Spektrum und dem nächsten Spektrum enthält. Damit werden stationäre Anteile in einem Signal, also zum Beispiel Signalanteile aufgrund von Harmonic Sustained Instrumenten robust und zuverlässig eliminiert. Dies liegt daran, dass die Differenzierung Änderungen in dem Signal hervorhebt und gleiche Anteile unterdrückt. So zeichnen sich jedoch perkussive Instrumente gerade dadurch aus, dass die durch diese Instrumente erzeugten Töne im Hinblick auf ihren Zeitverlauf stark transient sind.

[0053] Darüber hinaus wird es bevorzugt, die PCA 18a und die nicht-negative ICA 18b, also allgemeiner gesagt, die Zerlegungsoperation zum Zerlegen der extrahierten Kurzzeitspektren im Block 18 von Fig. 1 nicht mit den ursprünglichen Kurzzeitspektren sondern mit den abgeleiteten Kurzzeitspektren durchzuführen. Dabei wird der Effekt ausgenutzt, dass für stark transiente Signale das differenzierte Signal zum ursprünglichen Signal vor der Differenzierung sehr ähnlich ist, was insbesondere dann der Fall ist, wenn sehr schnelle Änderungen in einem Signal vorliegen. Dies gilt für perkussive Instrumente.

[0054] Des weiteren sei darauf hingewiesen, dass die Einrichtung 18 zum Zerlegen, die eine PCA 18a mit nachfolgender nichtnegativer ICA (18b) durchführt, ohnehin zur Ermittlung eines Profilspektrums eine gewichtete Linearkompensation der extrahierten Spektren, die von der Einrichtung geliefert werden, durchführt. Dies bedeutet, dass die extrahierten Spektren insgesamt mit bestimmten nach den einzelnen Verfahren berechneten Gewichtungsfaktoren beaufschlagt werden und linear kombiniert werden, also durch Subtraktion oder Addition. Daher wird zumindest teilweise der Effekt beobachtet, dass die Einrichtung 18 zum Hinterlegen der extrahierten Kurzzeitspektren eine der Differenzierung entgegenwirkende Funktionalität haben kann, so dass die Profilspektren, die für die Tonquellen ermittelt werden, nicht differenzierte Profilspektren sondern die eigentlichen Profilspektren sind. Auf jeden Fall hat sich herausgestellt, dass die Verwendung von differenzierten Spektren, also von Differenzen-Spektren aus einem Differenzen-Spektrogramm in Verbindung mit einem Zerlegungsalgorithmus in der Einrichtung 18, der auf einer gewichteten Linearkombination der einzelnen extrahierten Spektren basiert, zu Profilspektren für die einzelnen Tonquellen hoher Qualität und hoher Selektivität führen.

[0055] Würden dagegen lediglich stationäre Anteile weiter verarbeitet, also ist das spezifische Charakteristi-

kum nicht ein perkussives sondern ein harmonisches Charakteristikum, so wird es bevorzugt, eine Vorverarbeitung des Spektrogramms durch Integration, also durch Aufsummation zu erreichen, um die stationären Anteile gegenüber den transienten Anteilen zu verstärken. Auch in diesem Fall wird es bevorzugt, unter Verwendung der Summenspektren, also des integrierten Spektrogramms die Profilspektren für die einzelnen - dann harmonischen - Tonquellen zu berechnen.

[0056] Nachfolgend werden einzelne Funktionalitäten des erfindungsgemäßen Konzepts detaillierter dargestellt. Typische digitale Audiosignale werden jedoch bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung zunächst noch durch eine Vorverarbeitung der Einrichtung 8 vorverarbeitet. Ferner wird es bevorzugt, als PCM-Audiosignal, das in die Vorverarbeitungseinrichtung 8 eingegeben wird, Mono-Dateien mit einer Breite von 16 Bits pro Abtastwert bei einer Abtastfrequenz von 44,1 Hz zuzuführen. Diese Audiosignale, also dieser Strom von Audioabtastwerten, welcher auch ein Strom von Videoabtastwerten und allgemein ein Strom von Informations-Abtastwerten sein kann, wird der Vorverarbeitungseinrichtung 8 zugeführt, um eine Vorverarbeitung im Zeitbereich unter Verwendung einer Software-basierten Emulation eines Akustikeffektgeräts, das oft als "Exciter" bezeichnet wird, durchzuführen. Bei diesem Konzept verstärkt die Vorverarbeitungsstufe 8 den hochfrequenten Anteil des Audiosignals. Dies wird erreicht, in dem eine nicht-lineare Verzerrung mit einer hochpassgefilterten Version des Signals durchgeführt wird, und indem das Ergebnis der Verzerrung zu dem ursprünglichen Signal hinzuaddiert wird. Es stellt sich heraus dass diese Vorverarbeitung besonders günstig ist, wenn Hi-Hats zu beurteilen sind, oder ähnlich hoch klingende Idiophone mit niedriger Intensität. Ihr energetisches Gewicht bezüglich des Gesamtmusiksignals wird durch diesen Schritt erhöht, während die meisten harmonisch-ausgehaltenen Instrumente und Schlaginstrumente mit niedrigerem Ton nicht beeinträchtigt werden.

[0057] Ein weiterer positiver Seiteneffekt besteht in der Tatsache, dass MP3-kodierte und wieder dekodierte Dateien, die durch diesen Prozess inhärent tiefpassgefiltert wurden, wieder Hochfrequenzinformationen erhalten.

[0058] Eine Spektraldarstellung des vorverarbeiteten Zeitsignals wird dann unter Verwendung der Zeit/Frequenz-Einrichtung 12 erhalten, die vorzugsweise eine Kurzzeit-Fourier-Transformation (STFT; STFT = Short Time Fourier Transform) durchführt.

[0059] Zur Implementierung der Zeit/Frequenz-Einrichtung werden eine relativ große Blockgröße von vorzugsweise 4096 Werten und eine hohe Überlappung bevorzugt. Zunächst wird eine gute spektrale Auflösung für den niedrigeren Frequenzbereich, also für den niedrigeren Spektralkoeffizienten benötigt. Ferner wird die zeitliche Auflösung auf eine gewünschte Genauigkeit erhöht, indem eine kleine Hop-Größe, also ein kleines Hop-Intervall zwischen benachbarten Blöcken erhalten wird. Im bevorzugten Ausführungsbeispiel werden, wie es ausgeführt worden ist, 4096 Samples pro Block einer Kurzzeit-Fourier-Transformation unterworfen, was einer zeitlichen Blocklänge von 92 ms entspricht. Als Hop-Größe wird ein Wert von 10 ms verwendet. Dies bedeutet, dass jeder Abtastwert über 9 mal hintereinander in einem Kurzzeitspektren auftritt.

[0060] Die Einrichtung 12 ist ausgebildet, um ein Amplitudenspektrum X zu erhalten. Die Phaseninformationen können ebenfalls berechnet werden und, wie später noch ausgeführt wird, im Extremwert- bzw. Maximum-Sucher 16c verwendet werden.

[0061] Das Betragsspektrum X besitzt nunmehr n Frequenz-Bins oder Frequenz-Koeffizienten und m Spalten bzw. Rahmen (Frames), also einzelne Kurzzeitspektren. Die zeitvarianten Änderungen jedes Spektralkoeffizienten werden über allen Rahmen bzw. Einzelspektren differenziert, und zwar durch den Differenzierer 16a, um den Einfluss von harmonischausgehaltenen Tonquellen zu dezimieren, und um die nachfolgende Erfassung von Transienten zu vereinfachen. Die Differenzierung, die vorzugsweise eine Differenzbildung zwischen zwei Kurzzeitspektren der Folge aufweist, kann ferner noch gewisse Normierungen aufweisen.

[0062] Es sei darauf hingewiesen, dass die Differenzierung zu negativen Werten führen kann, so dass in einem Block 16b eine Halbwellengleichrichtung durchgeführt wird, um diesen Effekt zu entfernen. Alternativ könnten jedoch auch die negativen Vorzeichen einfach umgedreht werden, was jedoch im Hinblick auf die spätere Komponentenzerlegung nicht bevorzugt wird.

[0063] Aufgrund des Gleichrichters 16b wird somit ein nichtnegatives Differenz-Spektrogramm erhalten, das dem Maximum-Sucher 16c zugeführt wird.

[0064] Der Maximum-Sucher 16c führt eine Ereignis-Detektion durch, auf die nachfolgend eingegangen wird. Die Erfassung von mehreren lokalen Extremwerten und vorzugsweise von lokalen Maxima, die transienten Einsatz-Ereignissen in dem Musik-signal zugeordnet sind, wird durchgeführt, indem zunächst eine Zeittoleranz definiert wird, die zwei aufeinanderfolgende Schlagzeug-Einsätze trennt. Bei dem bevorzugten Ausführungsbeispiel wird eine Zeit von 68 ms als konstanter Wert verwendet, der von der Zeitauflösung und von Kenntnissen über das Musiksignal abgeleitet ist. Insbesondere bestimmt dieser Wert die Anzahl von Rahmen bzw. Einzelspektren bzw. differenzierten Einzelspektren, die zumindest zwischen zwei aufeinanderfolgenden Einsätzen auftreten muss. Die Verwendung dieses Minimalabstands wird auch durch die Betrachtung unterstützt, das eine Sechzehntelnote 60 ms bei einer oberen Tempogrenze von einem sehr hohen Tempo von 250 bpm dauert.

[0065] Um eine automatisierte Maximumsuche durchführen zu können, wird von dem differenzierten und gleichgerichteten Spektrum, also von der Folge von gleichgerichteten (differenten) Kurzzeitspektren eine Detektions-Funktion abgeleitet, auf Basis derer die Maximumsuche durchgeführt werden kann. Um für jeden

Zeitpunkt einen Wert dieser Funktion zu bekommen, wird einfach eine Summe über alle Frequenzkoeffizienten bzw. alle Spektral-Bins ermittelt. Zur Glättung dieser sich dann ergebenden eindimensionalen Funktion über der Zeit wird eine Faltung der erhaltenen Funktion mit einem geeigneten Hann-Fenster durchgeführt, so dass eine relativ glatte Funktion e erhalten wird. Um die Positionen t der Maxima zu erhalten, wird ein Gleitfenster der Toleranzlänge über den gesamten Weg e "geschoben", um die Fähigkeit zu erreichen, ein Maximum pro Schritt zu erhalten.

[0066] Die Verlässlichkeit der Maximasuche wird dadurch verbessert, dass vorzugsweise nur die Maxima beibehalten werden, die in einem Fenster für mehr als einen Zeitpunkt erscheinen, da sie sehr wahrscheinlich die interessierenden Peaks sind. So wird es bevorzugt, die Maxima zu verwenden, die über eine vorbestimmte Schwelle von Zeitpunkten, also zum Beispiel drei Zeitpunkte, ein Maximum darstellen, wobei die Schwelle letztendlich vom Verhältnis der Blocklänge zur Hop-Größe abhängen wird. Daraus ist ersichtlich, dass ein Maximum, wenn es wirklich ein signifikantes Maximum ist, eine bestimmte Anzahl von Zeitpunkten, also letztendlich eine bestimmte Anzahl von überlappenden Spektren tatsächlich ein Maximum sein muss, wenn daran gedacht wird, dass bei den vorher dargestellten Zahlenwerten jeder Abtastwert in wenigstens 9 aufeinanderfolgenden Kurzzeitspektren "mitmischt".

[0067] Bei dem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung werden, wie es durch den Phasenpfeil dargestellt ist, die aufgewickelten ("unwrapped") Phaseninformationen des ursprünglichen Spektrogramms als Zuverlässigkeitsfunktion verwendet. Es hat sich herausgestellt, dass in den Phaseninformationen ein signifikanter positiv gerichteter Phasensprung neben einer geschätzten Einsatz-Zeit t auftreten muss, wodurch vermieden wird, dass kleine Rippel fälschlicherweise als Einsätze bzw. "Onsets" betrachtet werden.

[0068] Erfindungsgemäß wird nunmehr ein kleiner Ausschnitt des Differenzspektrogramms, nämlich ein durch Differenzierung entstandenes Kurzzeitspektrum extrahiert und der nachfolgenden Zerlegungseinrichtung zugeführt.

[0069] Nachfolgend wird auf die Funktionalität der Einrichtung 18a zum Durchführen einer Principal-Component-Analysis eingegangen. Aus den in dem vorhergehenden Abschnitt beschriebenen Schritten werden also die Informationen über die Zeit des Auftretens **t** und die Spektralzusammensetzungen der Einsätze, also die extrahierten Kurzzeitspektren $\mathbf{X}_t$, abgeleitet. Bei reellen Musiksignalen findet man typischerweise eine große Anzahl von transienten Ereignissen innerhalb der Dauer des Musikstücks. Selbst bei einem einfachen Beispiel eines Stücks mit einer Geschwindigkeit von 120 Schlägen pro Minute (bpm) zeigt sich, dass in einem Vier-Minuten-Ausschnitt 480 Ereignisse sein können, gesetzt den Fall, dass nur Viertelnoten auftreten. Bezüglich des Ziels des Findens von nur ein paar wenigen signifikanten Unterräumen bzw. Profilspektren wird die Prinzipal-Komponenten-Analyse (PCA) auf $\hat{X}_t$ also auf die extrahierten Kurzzeitspektren oder auf von den extrahierten Kurzzeitspektren abgeleitete Kurzzeitspektren angewendet.

[0070] Unter Verwendung dieser bekannten Technik ist es möglich, der gesamte Satz von gesammelten Kurzzeitspektren auf eine begrenzte Anzahl von dekorrelierten Prinzipalkomponenten zu reduzieren, was in einer guten Darstellung der ursprünglichen Daten mit kleinem Rekonstruktionsfehler resultiert. Zu diesem Zweck wird eine Eigenwert-Zerlegung (EVD) der Kovarianz-Matrix des Datensatzes berechnet. Aus dem Satz von Eigenvektoren werden die Eigenvektoren mit den d größten Eigenwerten ausgewählt, um die Koeffizienten für die Linearkombination der ursprünglichen Vektoren gemäß der nachfolgenden Gleichung zu liefern:

$$\tilde{\mathbf{X}} = \hat{\mathbf{X}}_t \cdot \mathbf{T}$$

[0071] Daher beschreibt T eine Transformationsmatrix, die tatsächlich ein Teilsatz der Manigfaltigkeit der Eigenvektoren ist. Zusätzlich werden die reziproken Werte der Eigenwerte als Skalierungsfaktoren verwendet, was nicht nur zu einer Dekorrelation führt, sondern was auch eine Varianznormierung liefert, die wiederum zu einer Weißmachung bzw. einem Whitening-Effekt führt. Alternativ kann auch eine Singularwertzerlegung (SVD) von $\hat{X}_t$ verwendet werden. Es hat sich herausgestellt, dass die SVD äquivalent zur PCA mit EVD ist. Die weiß gemachten Komponenten $\tilde{X}$ werden nachfolgend in die ICA-Stufe 18b eingespeist, auf die nachfolgend eingegangen wird.

[0072] Allgemein gesagt ist die Independent-Component-Analysis (ICA) eine Technik, die verwendet wird, um einen Satz von linearen Mischsignalen in ihre ursprünglichen Quellen oder Komponentensignale zu zerlegen. Eine Anforderung für ein optimales Verhalten des Algorithmus ist die statistische Unabhängigkeit der Quellen. Vorzugsweise wird eine nicht-negative ICA verwendet, die auf dem intuitiven Konzept des Optimierens einer Kostenfunktion aufbaut, die die Nicht-Negativität der Komponenten beschreibt. Diese Kostenfunktion ist auf einen Rekonstruktionsfehler bezogen, der durch Achsenpaarrotationen von zwei oder mehr Variablen in dem positiven Quadranten der gemeinsamen Wahrscheinlichkeitsdichtefunktion (PDF) eingeführt wird. Die Annahmen für dieses Modell implizieren, dass die ursprünglichen Quellensignale positiv sind und bei Null eine PDF ungleich Null haben, und dass sie bis zu einem gewissen Grad linear unabhängig sind. Das erste Konzept wird immer erfüllt, da die Vektoren, die der ICA unterzogen werden, aus der differenzierten und halbwellen-gleichgewichteten Version $\hat{X}$ des ursprünglichen Spektro-

gramms **X** resultieren, welche somit niemals Werte kleiner als Null umfasst, jedoch sicherlich Werte gleich Null. Die zweite Begrenzung wird berücksichtigt, wenn die zu Einsatzzeiten gesammelten Spektren als die Linearkombinationen eines kleinen Satzes von ursprünglichen Quellenspektren betrachtet werden, die die betrachteten Instrumente charakterisieren. Dies bedeutet natürlich eine ziemlich grobe Annäherung, sie stellt sich jedoch in der Vielzahl der Fälle als ausreichend gut heraus.

[0073] Ferner wird vorzugsweise davon Gebrauch gemacht, dass die Spektren, die Einsätze haben, und zwar insbesondere die Spektren von tatsächlichen Schlagzeuginstrumenten keine invarianten Strukturen haben, sondern im Hinblick auf ihre spektrale Zusammensetzung hier keinen Änderungen unterzogen werden. Nichtsdestoweniger kann jedoch angenommen werden, dass es charakteristische Eigenschaften gibt, die für Spektralprofile von Schlagzeugtönen charakteristisch sind, die es somit ermöglichen, dass die weiß gemachten Komponenten $\tilde{X}$ in ihre potentielle Quellen- bzw. Profilspektren F gemäß der nachfolgenden Gleichung aufgetrennt werden.

$$\mathbf{F} = \mathbf{A} \cdot \tilde{\mathbf{X}}$$

[0074] A bezeichnet eine d x d Entmisch-Matrix, die durch den ICA-Prozess bestimmt wird, der tatsächlich die einzelnen Komponenten $\tilde{X}$ trennt. Die Quellen F werden in diesem Dokument auch als Profilspektren bezeichnet. Jedes Profilspektrum hat genauso wie ein Spektrum des ursprünglichen Spektrogramms n Frequenzbins, ist jedoch für alle Zeiten - abgesehen von der Amplitudennormierung - also die Amplitudenhüllkurve - identisch. Dies bedeutet, dass ein solches Profilspektrum nur die Spektralinformationen enthält, die auf ein Onset-Spektrum eines Instruments bezogen sind. Um eine beliebige Skalierung der Komponenten, die durch PCA und ICA eingeführt werden, vorzugsweise zu umgehen, wird eine Transformationsmatrix R gemäß folgender Gleichung verwendet:

$$\mathbf{R} = \mathbf{T} \cdot \mathbf{A}^T$$

[0075] Die Normierung von R mit seinem absoluten Maximalwert führt zu Gewichtungskoeffizienten in einem Bereich von -1 bis +1, so dass Spektralprofile, die unter Verwendung der nachfolgenden Gleichung extrahiert werden

$$\mathbf{F} = \hat{\mathbf{X}}_t \cdot \mathbf{R}$$

[0076] Werte in dem Bereich des ursprünglichen Spektrogramms haben. Eine weitere Normierung wird durch Teilen jedes Spektralprofils durch seine L2-Norm erreicht.

[0077] Wie es bereits vorher ausgeführt worden ist, ist die Annahme der Unabhängigkeit und die Annahme der Invarianz für gegebene Kurzzeitspektren nicht immer hundertprozentig erfüllt. So ist es keine Überraschung, dass die nach der Entmischung erhaltenen Spektralprofile immer noch bestimmte Abhängigkeiten haben können. Dies sollte jedoch nicht als fehlerhaftes Verhalten betrachtet werden. Tests mit Spektralprofilen von einzelnen Schlagzeugtönen haben gezeigt, dass die Spektralprofile ebenfalls eine starke Abhängigkeit zwischen den Einsatz-Spektren unterschiedlicher perkussiver Instrumente haben. Eine Art und Weise zum Messen des Grads an gegenseitiger Überlappung und Ähnlichkeit entlang der Frequenzachse besteht in der Durchführung von Crosstalk-Messungen. Aus Anschauungsgründen können die Spektralprofile, die von dem ICA-Prozess erhalten werden, als Übertragungsfunktion von stark frequenzselektiven Teilen in einer Filterbank betrachtet werden, wobei überlappenden Durchgangsbänder zu Crosstalk in dem Ausgang der Filterbankkanäle führen kann. Das Crosstalkmaß zwischen zwei Spektralprofilen wird gemäß folgender Gleichung berechnet.

$$\mathbf{C}_{i,j} = \frac{\mathbf{F}_i \cdot \mathbf{F}_j^T}{\mathbf{F}_i \cdot \mathbf{F}_i^T}$$

[0078] In der vorstehenden Gleichung reicht i von 1 bis d, reicht j von 1 bis d und gilt, dass j ungleich i ist. In der Tat ist dieser Wert auf den bekannten Kreuzkorrelationskoeffizienten bezogen, derselbe verwendet jedoch eine andere Normierung.

[0079] Basierend auf den bestimmten Profilspektren wird nunmehr im Block 20 von Fig. 2 eine Amplitudenhüllkurvenbestimmung durchgeführt. Hierzu wird das ursprüngliche Spektrogramm, also die Folge von z.B. durch die Einrichtung 12 von Fig. 1 oder in Zeit/Frequenz/Umsetzer 12 von Fig. 2 erhaltenen Kurzzeitspektren verwendet. Folgende Gleichung gilt:

$$\mathbf{E} = \mathbf{F} \cdot \mathbf{X}$$

[0080] Als zweite Informationsquelle kann auch die differenzierte Version der Amplitudenhüllkurven aus dem Differenzspektrogramm gemäß folgender Gleichung ermittelt werden:

$$\hat{\mathbf{E}} = \mathbf{F} \cdot \hat{\mathbf{X}}$$

[0081] Wesentlich an diesem Konzept ist, dass keine weitere ICA-Berechnung mit den Amplitudenhüllkurven durchgeführt wird. Stattdessen werden durch das erfindungsgemäße Konzept hoch spezialisierte Spektralprofile erhalten, die sehr nahe an den Spektren der Instrumente liegen, die tatsächlich in dem Signal erscheinen. Dennoch sind die extrahierten Amplitudenhüllkurven nur in bestimmten Fällen schöne Erfassungsfunktionen mit scharfen Spitzen, beispielsweise für tanzorientierte Musik mit sehr dominierenden perkussiven Rhythmusanteilen. Oft enthalten die Amplitudenhüllkurven kleinere Spitzen und Plateaus, die von den oben erwähnten Crosstalk-Effekten herrühren können.

[0082] Nachfolgend wird auf eine nähere Implementierung der Einrichtung 22 zur Merkmalsextraktion und Klassifikation hingewiesen. Es ist bekannt, dass die tatsächliche Anzahl von Komponenten für reelle Musiksignale zunächst einmal unbekannt ist. "Komponenten" bedeuten in diesem Kontext sowohl die Spektralprofile als auch die korrespondierenden Amplitudenhüllkurven. Wenn die Anzahl d von extrahierten Komponenten zu niedrig ist, werden Artefakte der nicht berücksichtigten Komponenten sehr wahrscheinlich in anderen Komponenten auftreten. Wenn dagegen zu viele Komponenten extrahiert werden, sind die prominentesten Komponenten in mehre Komponenten aufgeteilt. Ungünstigerweise kann diese Aufteilung selbst mit der richtigen Anzahl von Komponenten auftreten und gelegentlich eine Erfassung der reellen Komponenten erschweren.

[0083] Zur Überwindung dieser Problematik wird eine maximale Anzahl d von Komponenten in dem PCA- oder ICA-Prozess vorgegeben. Anschließend werden die extrahierten Komponenten unter Verwendung eines Satzes von spektralbasierten und zeitbasierten Merkmalen klassifiziert. Die Klassifizierung soll zwei Informationen liefern. Zunächst sollen die Komponenten aus dem weiteren Verfahren eliminiert werden, die mit hoher Sicherheit als nicht-perkussiv erkannt werden. Ferner sollen die verbleibenden Komponenten vordefinierten Instrumentenklassen zugeordnet werden.

[0084] Ein geeignetes Maß für die Unterscheidung der Amplitudenhüllkurven wird durch die Perkussivität gegeben, die in der dritten Fachveröffentlichung genannt ist. Hier wird eine modifizierte Version verwendet, bei der der Korrelationskoeffizient zwischen entsprechenden Amplitudenhüllkurven in $\hat{E}$ und E verwendet wird. Der Grad an Korrelation zwischen beiden Vektoren tendiert dazu, klein zu sein, wenn die charakteristischen Plateaus, die auf harmonisch ausgehaltene Töne bezogen sind, in den nicht-differenzierten Amplitudenhüllkurven E auftauchen. Diese verschwinden sehr wahrscheinlich in der differenzierten Version $\hat{E}$. Beide Vektoren sind sich im Falle von transienten Amplitudenhüllkurven, die von perkussiven Tönen stammen, wesentlich ähnlicher. Zu diesem Zweck wird auf Fig. 3a und Fig. 4a verwiesen. In Fig. 3a ist eine sehr schnell und sehr hoch ansteigende Amplitudenhüllkurve für eine perkussive Quelle gezeigt, während in Fig. 4a eine Amplitudenhüllkurve für ein harmonisch ausgehaltenes Instrument gezeigt ist. Fig. 3a ist eine Amplitudenhüllkurve für eine Kick Drum, während Fig. 4a eine Amplitudenhüllkurve für eine Trompete ist. Aus der Amplitudenhüllkurve für die Trompete ist ein relativ zügiger Anstieg, und dann aber ein relativ langsames Ausklingen dargestellt, wie es für harmonisch ausgehaltene Instrumente typisch ist. Dagegen steigt die Amplitudenhüllkurve für ein perkussives Element, wie es in Fig. 3a gezeigt ist, sehr schnell und sehr stark an und fällt jedoch ebenfalls wieder genauso schnell und steil ab, da ein Schlagzeugton typischerweise aufgrund des Wesens der Erzeugung dieses Tons nicht besonders lang nachklingt bzw. abklingt.

[0085] Die Amplitudenhüllkurven können somit zur Klassifikation bzw. Merkmalsextraktion genauso gut verwendet werden, wie die nachfolgend erläuterten Profilspektren, die sich im Falle einer perkussiven Quelle (Fig. 3b; Hi-Hat) und Fig. 4b im Falle eines harmonisch ausgehaltenen Instruments (Gitarre) deutlich unterscheiden. So ist beim harmonisch ausgehaltenen Instrument eine deutliche Ausprägung der Oberwellen zu sehen, während die perkussive Quelle ein eher rauschartiges Spektrum hat, das keine deutlich ausgeprägten Oberwellen hat, das jedoch insgesamt einen Bereich hat, in dem Energie konzentriert ist, wobei dieser Bereich, in dem Energie konzentriert ist, sehr breitbandig ist.

[0086] Es wird also vorzugsweise ein spektral-basiertes Maß, also ein Maß, das von den Profilspektren (z.B. Fig. 3b und Fig. 4b) abgeleitet wird, verwendet, um Spektren von harmonisch ausgehaltenen Tönen von Spektren, die auf perkussive Töne bezogen sind, zu trennen. Wieder wird bei dem bevorzugten Ausführungsbeispiel eine modifizierte Version der Berechnung dieses Maßes verwendet, die eine Toleranz gegenüber spektralen Lag-Erscheinungen, eine Dissonanz mit allen Harmonischen und einer geeigneten Normierung zeigt. Ein höherer Grad an rechenmäßiger Effizienz wird erreicht, indem eine ursprüngliche Dissonanzfunktion mit einer Gewichtungsmatrix für Frequenzpaare ersetzt wird.

[0087] Die Zuordnung von spektralen Profilen zu aprioridefinierten Klassen von perkussiven Instrumenten wird durch einen einfachen Klassifizierer zum Klassifizieren der k nächsten Nachbarn mit Spektralprofilen von einzelnen Instrumenten als Trainingsdatenbank geschaffen. Die Distanzfunktion wird aus wenigstens einem Korrelationskoeffizient zwischen einem Abfrageprofil und einem Datenbankprofil berechnet. Um die Klassifikation in Fällen niedriger Zuverlässigkeit, also bei niedrigen Korrelationskoeffizienten, zu verifizieren, oder um ein mehrmaliges Auftreten derselben Instrumente zu verifizieren, werden zusätzliche Merkmale, die eine detaillierte Information über die Form des Spektralprofils liefern, extrahiert. Diese umfassen die bereits vorher ge-

nannten einzelnen Merkmale.

[0088] Nachfolgend wird weiter auf die Funktionalität des Entscheiders 24 in Fig. 2 eingegangen. Schlagzeugartige Einsätze werden in den Amplitudenhüllkurven, wie beispielsweise in der Amplitudenhüllkurve in Fig. 3a, unter Verwendung üblicher Spitzenauswahlverfahren, die auch als Peak-Picking bezeichnet sind, erfasst. Nur Spitzen in einem Toleranzbereich neben den ursprünglichen Zeiten t, also den Zeiten, in denen der Maximumsucher 16c ein Ergebnis lieferte, werden vordringlich als Kandidaten für Einsätze betrachtet. Restliche aus den Amplitudenhüllkurven extrahierte Spitzen werden für weitere Betrachtungen zunächst gespeichert. Der Wert des Betrags der Amplitudenhüllkurve wird jedem Einsatz-Kandidat an seiner Position zugeordnet. Wenn dieser Wert nicht einen vorbestimmten dynamischen Schwellenwert überschreitet, dann wird der Einsatz nicht akzeptiert. Die Schwelle variiert über der Menge an Energie in einem größeren zeitlichen Bereich, der die Einsätze umgibt. Der größte Teil des Crosstalk-Einflusses von harmonisch ausgehaltenen Instrumenten sowie gleichzeitig spielenden perkussiven Instrumenten kann in diesem Schritt reduziert werden. Ferner wird es bevorzugt, zu unterscheiden, ob gleichzeitige Einsätze von unterschiedlichen perkussiven Instrumenten tatsächlich vorhanden sind oder nur aufgrund von Crosstalk-Effekten existieren. Eine Lösung für dieses Problem besteht vorzugsweise darin, diese weiteren Auftrittsereignisse zu akzeptieren, deren Wert im Vergleich zum Wert des stärksten Instruments zum Einsatzzeitpunkt relativ hoch ist.

[0089] Erfindungsgemäß wird somit eine automatische Erfassung und vorzugsweise auch eine automatische Klassifikation von nicht-gepitchten perkussiven Instrumenten in reellen polyphonen Musiksignalen erreicht, wobei die Ausganqsbasis hierfür die Profilspektren einerseits und die Amplitudenhüllkurve andererseits sind. Aus den perkussiven Instrumenten kann ferner gut die rhythmische Information eines Musikstücks extrahiert werden, was wiederum zu einer günstigen Noten-zu-Noten-Transkription führen dürfte.

[0090] Abhängig von den Gegebenheiten kann das erfindungsgemäße Verfahren zum Analysieren eines Informationssignals in Hardware oder in Software implementiert werden. Die Implementierung kann auf einen 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 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 Analysieren eines Informationssignals, mit folgenden Merkmalen:

   einer Einrichtung (16) zum Extrahieren von signifikanten Kurzzeitspektren oder von Kurzzeitspektren des Informationssignals abgeleiteten signifikanten Kurzzeitspektren aus einer zeitlichen Folge von Kurzzeitspektren, die das Informationssignal darstellen, wobei die Einrichtung (16) zum Extrahieren ausgebildet ist, um solche Kurzzeitspektren aus der zeitlichen Folge zu extrahieren, die einer spezifischen Charakteristik näher kommen als andere Kurzzeitspektren des Informationssignals;
   einer Einrichtung (18) zum Zerlegen der extrahierten Kurzzeitspektren in Komponentensignalspektren, wobei ein Komponentensignalspektrum ein Profilspektrum einer Tonquelle darstellt, die einen Ton erzeugt, der der gesuchten Charakteristik entspricht, und wobei ein anderes Komponentensignalspektrum ein Profilspektrum einer anderen Tonquelle darstellt, die einen Ton erzeugt, der der gesuchten Charakteristik entspricht; und
   einer Einrichtung (20) zum Berechnen einer Amplitudenhüllkurve für die Tonquellen, wobei eine Amplitudenhüllkurve für eine Tonquelle angibt, wie sich ein Profilspektrum der Tonquelle über der Zeit ändert, unter Verwendung der Profilspektren und einer Folge von Kurzzeitspektren, die das Informationssignal darstellt.

2. Vorrichtung nach Anspruch 1, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um das Informationssignal derart vorzuverarbeiten (8), dass Signalanteile in dem Informationssignal bei höheren Frequenzen gegenüber Signalanteilen in dem Informationssignal bei niedrigeren Frequenzen in dem Informationssignal hervorgehoben werden.

3. Vorrichtung nach Anspruch 2, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um bei der Vorverarbeitung (8)
   das Informationssignal einer Hochpassfilterung zu unterziehen,
   die hochpassgefilterte Version des Informationssignals nicht-linear zu verzerren, und
   das nicht-linear verzerrte Signal zu dem ursprünglichen Informationssignal hinzuzuaddieren.

4. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um das Informationssignal einer Zeitbereich-Frequenzbereich-Konversion (12) zu unterziehen, um eine Folge von Kurzzeitspektren zu erhalten, wobei sich zwei zeitlich benachbarte Kurz-

zeitspektren auf Ausschnitte des Informationssignals beziehen, die sich bis auf ein Hopping-Intervall überlappen.

5. Vorrichtung nach Anspruch 4, bei der jedes Kurzzeitspektrum jeweils eine Folge von Spektralkoeffizienten aufweist, und
bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um die Folge von Kurzzeitspektren in zeitlicher Hinsicht zu differenzieren (16a), um eine Folge von differenzierten Kurzzeitspektren zu erhalten, wobei ein differenziertes Kurzeitspektrum Informationen über Änderungen in einem Kurzzeitspektrum zu einem zeitlich vorhergehenden oder zeitlich nachfolgendem Kurzzeitspektrum nachweist.

6. Vorrichtung nach Anspruch 5, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um ein differenziertes Kurzzeitspektrum zu erhalten, indem für jeden Spektralkoeffizienten eine Differenz des Spektralkoeffizienten in einem aktuellen Kurzzeitspektrum und einem vorhergehenden oder nachfolgenden Kurzzeitspektrum gebildet wird.

7. Vorrichtung nach Anspruch 5 oder 6, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um die differenzierten Kurzzeitspektren gleichzurichten (16b), so dass ein gleichgerichtetes differenziertes Kurzzeitspektrum keine negativen Werte aufweist.

8. Vorrichtung nach einem der Ansprüche 5 bis 7, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um signifikante Kurzzeitspektren basierend auf den differenzierten Kurzzeitspektren zu ermitteln.

9. Vorrichtung nach Anspruch 8, wobei die Einrichtung (16) zum Extrahieren ausgebildet ist, um für jedes differenzierte Kurzzeitspektrum Spektralkoeffizienten oder von Spektralkoeffizienten abgeleitete Werte aus dem differenzierten Kurzzeitspektrum aufzusummieren (16c), um für ein Kurzzeitspektrum einen Summenwert zu erhalten, so dass sich eine Detektionsfunktion über der Zeit ergibt.

10. Vorrichtung nach Anspruch 9, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um die Detektionsfunktion über der Zeit zu glätten.

11. Vorrichtung nach Anspruch 9 oder 10, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um Maxima in der Detektionsfunktion zu einem Zeitpunkt zu finden (16c), und um ein differenziertes Kurzzeitspektrum oder ein Kurzzeitspektrum als signifikantes Spektrum zu verwenden, dem ein Zeitpunkt zugeordnet ist, an dem die Detektionsfunktion ein Maximum aufweist.

12. Vorrichtung nach einem der Ansprüche 9 bis 11, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um nur Maxima der Detektionsfunktion als signifikant zu erachten, die mehr als eine vordefinierte Zeitspanne voneinander zeitlich beabstandet sind.

13. Vorrichtung nach einem der Ansprüche 4 bis 12, bei der die Einrichtung (16) zum Extrahieren ausgebildet ist, um als Folge von Kurzzeitspektren Betragsspektren zu ermitteln und Phaseninformationen der Kurzzeitspektren bei der Extraktion der signifikanten Kurzzeitspektren zu verwenden.

14. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (18) zum Zerlegen ausgebildet ist, um die extrahierten Kurzzeitspektren gewichtet zu addieren (18a), um eine reduzierte Anzahl von extrahierten Kurzzeitspektren zu erhalten.

15. Vorrichtung nach einem der Ansprüche 1 bis 14, bei der die Einrichtung (18) zum Zerlegen ausgebildet ist, um zur Dimensionsreduktion eine Prinzipal-Komponenten-Analyse durchzuführen (18a), um verarbeitete Kurzzeitspektren zu erhalten.

16. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (18) zum Zerlegen ausgebildet ist, um eine Independent-Component-Analysis (18b) durchzuführen, um eine Mehrzahl von Komponentensignalen zu erzeugen, wobei einem Komponentensignal eine Informationsquelle, die zu dem Informationssignal beiträgt, zugeordnet ist.

17. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (20) zum Berechnen der Amplitudenhüllkurve ausgebildet ist, um eine Matrix, die die Profilspektren umfasst, und eine Matrix, die eine Folge von Kurzzeitspektren des Informationssignals umfasst, zu multiplizieren, um für die Tonquellen die Amplitudenhüllkurven zu erhalten.

18. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung zum Berechnen der Amplitudenhüllkurve ausgebildet ist, um ferner eine differenzierte Amplitudenhüllkurve unter Verwendung der Profilspektren für die Tonquellen und unter Verwendung des Differenzspektrogramms zu ermitteln.

19. Vorrichtung nach einem der vorhergehenden Ansprüche, die ferner eine Einrichtung (22) zum Klassifizieren der Komponentensignale in perkussive Komponentensignale und nicht-perkussive Komponentensignale umfasst.

20. Vorrichtung nach Anspruch 19, bei der die Einrichtung (22) zum Klassifizieren ausgebildet ist, um auf

der Basis der Profilspektren und/oder der Amplitudenhüllkurven zu klassifizieren.

21. Vorrichtung nach Anspruch 19 oder 20, bei der die Einrichtung (20) zum Klassifizieren ausgebildet ist, um aus den Profilspektren oder den Amplitudenhüllkurven ein Merkmal zu extrahieren und mit Merkmalen bekannter Quellen in einer Datenbank zu vergleichen.

22. Vorrichtung nach einem der vorhergehenden Ansprüche, die ferner eine Einrichtung (24) zum Untersuchen der Amplitudenhüllkurven für eine Tonquelle aufweist, um ein Maximum in der Amplitudenhüllkurve dann als einen Einsatz eines Signals von der Tonquelle zu akzeptieren, wenn die Einrichtung (16) zum Extrahieren zu einem innerhalb einer Schwelle ähnlichen Zeitpunkt ein signifikantes Kurzzeitspektrum extrahiert hatte.

23. Vorrichtung nach einem der vorhergehenden Ansprüchen, bei dem die Einrichtung (20) zum Berechnen der Amplitudenhüllkurve ausgebildet ist, um die Amplitudenhüllkurve für eine Tonquelle so zu berechnen, dass die Amplitudenhüllkurve angibt, wie sich eine Intensität oder Gewichtung eines Profilspektrums der Tonquelle über der Zeit ändert.

24. Verfahren zum Analysieren eines Informationssignals, mit folgenden Schritten:

Extrahieren (16) von signifikanten Kurzzeitspektren oder von Kurzzeitspektren des Informationssignals abgeleiteten signifikanten Kurzzeitspektren aus einer zeitlichen Folge von Kurzzeitspektren, die das Informationssignal darstellen, wobei solche Kurzzeitspektren aus der zeitlichen Folge extrahiert werden, die einer spezifischen Charakteristik näher kommen als andere Kurzzeitspektren des Informationssignals;
Zerlegen (18) der extrahierten Kurzzeitspektren in Komponentensignalspektren, wobei ein Komponentensignalspektrum ein Profilspektrum einer Tonquelle darstellt, die einen Ton erzeugt, der der gesuchten Charakteristik entspricht, und wobei ein anderes Komponentensignalspektrum ein Profilspektrum einer anderen Tonquelle darstellt, die einen Ton erzeugt, der der gesuchten Charakteristik entspricht; und
Berechnen (20) einer Amplitudenhüllkurve für die Tonquellen, wobei eine Amplitudenhüllkurve für eine Tonquelle angibt, wie sich ein Profilspektrum der Tonquelle über der Zeit ändert, unter Verwendung der Profilspektren und einer Folge von Kurzzeitspektren, die das Informationssignal darstellt.

25. Computer-Programm mit einem Programmcode angepasst um ein Verfahren zum Analysieren eines Informations-signals gemäß Patentanspruch 24, durchzuführen, wenn das Computer-Programm auf einem Computer abläuft.

Claims

1. Device for analyzing an information signal, comprising:

means (16) for extracting significant short-time spectra or significant short-time spectra, which are derived from short-time spectra of the information signal, from a time sequence of short-time spectra representing the information signal, the means (16) for extracting being configured to extract, from the time sequence, such short-time spectra which come closer to a specific characteristic than other short-time spectra of the information signal;
means (18) for decomposing the extracted short-time spectra into component signal spectra, a component signal spectrum representing a profile spectrum of a tone source which generates a tone corresponding to the characteristic sought for, and another component signal spectrum representing a profile spectrum of another tone source which generates a tone corresponding to the characteristic sought for; and
means (20) for calculating an amplitude envelope for the tone sources, an amplitude envelope for a tone source indicating how a profile spectrum of the tone source changes over time, using the profile spectra and a sequence of short-time spectra representing the information signal.

2. Device as claimed in claim 1, wherein the means (16) for extracting is configured to pre-process (8) the information signal such that signal portions present in the information signal at higher frequencies are accentuated, in the information signal, as compared to signal portions present in the information signal at lower frequencies.

3. Device as claimed in claim 2, wherein the means (16) for extracting is configured to
subject the information signal to high-pass filtering,
distort the high-pass filtered version of the information signal in a non-linear manner, and
add the non-linearly distorted signal to the original information signal
in pre-processing (8).

4. Device as claimed in one of the preceding claims, wherein the means (16) for extracting is configured

to subject the information signal to a time-domain/frequency-domain conversion (12) to obtain a sequence of short-time spectra, two short-time spectra adjacent in time relating to portions of the information signal which overlap except for a hopping interval.

5. Device as claimed in claim 4, wherein each short-time spectrum comprises a sequence of spectral coefficients, and
wherein the means (16) for extracting is configured to differentiate (16a) the sequence of short-time spectra in terms of time to obtain a sequence of differentiated short-time spectra, a differentiated short-time spectrum proving information about changes in a short-time spectrum as compared to a preceding or subsequent short-time spectrum.

6. Device as claimed in claim 5, wherein the means (16) for extracting is configured to obtain a differentiated short-time spectrum by forming, for each spectral coefficient, a difference of the spectral coefficient in a current short-time spectrum and a previous or subsequent short-time spectrum.

7. Device as claimed in claim 5 or 6, wherein the means (16) for extracting is configured to rectify (16b) the differentiated short-time spectra so that a rectified differentiated short-time spectrum does not exhibit any negative values.

8. Device as claimed in one of claims 5 to 7, wherein the means (16) for extracting is configured to determine significant short-time spectra on the basis of the differentiated short-time spectra.

9. Device as claimed in claim 8, wherein the means (16) for extracting is configured to sum up (16c), for each differentiated short-time spectrum, spectral coefficients, or values derived from spectral coefficients, from the differentiated short-time spectrum, so as to obtain a sum value for a short-time spectrum, so that a detection function over time results.

10. Device as claimed in claim 9, wherein the means (16) for extracting is configured to smooth the detection function over time.

11. Device as claimed in claim 9 or 10, wherein the means (16) for extracting is configured to find (16c) maxima in the detection function at a point in time, and to use a differentiated short-time spectrum or a short-time spectrum as a significant spectrum having a point in time associated therewith at which the detection function exhibits a maximum.

12. Device as claimed in one of claims 9 to 11, wherein the means (16) for extracting is configured to regard only such maxima of the detection function as significant that are spaced apart in time by more than a predefined time period.

13. Device as claimed in one of claims 4 to 12, wherein the means (16) for extracting is configured to determine amount spectra as a sequence of short-time spectra, and to use phase information of the short-time spectra when extracting the significant short-time spectra.

14. Device as claimed in one of the preceding claims, wherein the means (18) for decomposing is configured to add (18a) the extracted short-time spectra in a weighted manner so as to obtain a reduced number of extracted short-time spectra.

15. Device as claimed in one of claims 1 to 14, wherein the means (18) for decomposing is configured to perform (18a) a principal component analysis for dimension reduction so as to obtain processed short-time spectra.

16. Device as claimed in one of the preceding claims, wherein the means (18) for decomposing is configured to perform an independent component analysis (18b) to produce a plurality of component signals, a component signal being associated with an information source contributing to the information signal.

17. Device as claimed in one of the preceding claims, wherein the means (20) for calculating the amplitude envelope is configured to multiply a matrix including the profile spectra and a matrix including a sequence of short-time spectra of the information signal, so as to obtain the amplitude envelopes for the tone sources.

18. Device as claimed in one of the preceding claims, wherein the means for calculating the amplitude envelope is configured to further determine a differentiated amplitude envelope using the profile spectra for the tone sources and using the difference spectrogram.

19. Device as claimed in one of the preceding claims, further comprising means (22) for classifying the component signals into percussive component signals and non-percussive component signals.

20. Device as claimed in claim 19, wherein the means (22) for classifying is configured to perform a classification on the basis of the profile spectra and/or the amplitude envelopes.

21. Device as claimed in claim 19 or 20, wherein the means (20) for classifying is configured to extract a feature from the profile spectra or the amplitude envelopes, and to compare it with features of known

sources in a database.

22. Device as claimed in one of the preceding claims, further comprising means (24) for examining the amplitude envelopes for a tone source to accept a maximum in the amplitude envelope as an onset of a signal from the tone source when the means (16) for extracting had extracted a significant short-time spectrum at a point in time similar within a threshold.

23. Device as claimed in one of the preceding claims, wherein the means (20) for calculating the amplitude envelope is configured to calculate the amplitude envelope for a tone source in such a manner that the amplitude envelope indicates how an intensity or weighting of a profile spectrum of the tone source changes over time.

24. Method for analyzing an information signal, comprising:

extracting (16) significant short-time spectra or significant short-time spectra, which are derived from short-time spectra of the information signal, from a time sequence of short-time spectra representing the information signal, the short-time spectra extracted from the time sequence being such short-time spectra which come closer to a specific characteristic than other short-time spectra of the information signal;
decomposing (18) the extracted short-time spectra into component signal spectra, a component signal spectrum representing a profile spectrum of a tone source which generates a tone corresponding to the characteristic sought for, and another component signal spectrum representing a profile spectrum of another tone source which generates a tone corresponding to the characteristic sought for; and
calculating (20) an amplitude envelope for the tone sources, an amplitude envelope for a tone source indicating how a profile spectrum of the tone source changes over time, using the profile spectra and a sequence of short-time spectra representing the information signal.

25. Computer program having a program code adapted to perform the method for analyzing an information signal according to claim 24, when the computer program runs on a computer.

Revendications

1. Dispositif pour analyser un signal d'information, aux caractéristiques suivantes:

un moyen (16) pour extraire des spectres de courte durée significatifs ou des spectres de courte durée significatifs dérivés de spectres de courte durée du signal d'information d'une succession dans le temps de spectres de courte durée représentant le signal d'information, le moyen (16) pour extraire étant réalisé de manière à extraire de la succession dans le temps les spectres de courte durée qui se rapprochent plus d'une caractéristique spécifique que d'autres spectres de courte durée du signal d'information;
un moyen (18) pour décomposer les spectres de courte durée extraits en spectres de signal à composantes, un spectre de signal à composantes représentant un spectre de profil d'une source sonore qui génère un son correspondant à la caractéristique recherchée, et un autre spectre de signal à composantes représentant un spectre de profil d'une autre source sonore qui génère un son correspondant à la caractéristique recherchée; et
un moyen (20) pour calculer une courbe enveloppante d'amplitude pour les sources sonores, une courbe enveloppante d'amplitude pour une source sonore indiquant la manière dont varie dans le temps un spectre de profil de la source sonore, à l'aide des spectres de profil et d'une succession de spectres de courte durée représentant le signal d'information.

2. Dispositif selon la revendication 1, dans lequel le moyen (16) pour extraire est réalisé de manière à prétraiter le signal d'information (8) de sorte que les parties de signal dans le signal d'information à hautes fréquences sont mises en évidence par rapport aux parties de signal dans le signal d'information à basses fréquences dans le signal d'information.

3. Dispositif selon la revendication 2, dans lequel le moyen (16) pour extraire est réalisé de manière à, lors du prétraitement (8),
soumettre le signal d'information, à une filtration passe-haut,
distorsionner de manière non linéaire la version filtrée passe-haut du signal d'information, et
ajouter le signal distorsionné non linéairement au signal d'information original.

4. Dispositif selon l'une des revendications précédentes, dans lequel le moyen (16) pour extraire est réalisé de manière à soumettre le signal d'information à une conversion domaine du temps-domaine de la fréquence (12), pour obtenir une succession de spectres de courte durée, deux spectres de courte durée adjacents dans le temps se rapportant à des segments du signal d'information qui se recouvrent à un intervalle de saut près.

5. Dispositif selon la revendication 4, dans lequel chaque spectre de courte durée présente une succession de coefficients spectraux, et
dans lequel le moyen (16) pour extraire est réalisé de manière à différentier la succession de spectres de courte durée en ce qui concerne le temps (16a), pour obtenir une succession de spectres de courte durée différentiés, un spectre de courte durée différentié donnant des informations sur des variations dans un spectre de courte durée par rapport à un spectre de courte durée précédent dans le temps ou suivant dans le temps.

6. Dispositif selon la revendication 5, dans lequel le moyen (16) pour extraire est réalisé de manière à obtenir un spectre de courte durée différentié en formant, pour chaque coefficient spectral, une différence du coefficient spectral dans un spectre de courte durée actuel et un spectre de courte durée précédent ou suivant.

7. Dispositif selon la revendication 5 ou 6, dans lequel le moyen (16) pour extraire est réalisé de manière à redresser les spectres de courte durée différentiés (16b) de sorte qu'un spectre de courte durée différentié redressé ne présente pas de valeurs négatives.

8. Dispositif selon l'une des revendications 5 à 7, dans lequel le moyen (16) pour extraire est réalisé de manière à déterminer les spectres de courte durée significatifs sur base des spectres de courte durée différentiés.

9. Dispositif selon la revendication 8, dans lequel le moyen (16) pour extraire est réalisé de manière à additionner, pour chaque spectre de courte durée différentié, des coefficients spectraux ou des valeurs dérivées des coefficients spectraux du spectre de courte durée différentié (16c), pour obtenir, pour un spectre de courte durée, une valeur de somme, de sorte qu'il s'ensuive une fonction de détection dans le temps.

10. Dispositif selon la revendication 9, dans lequel le moyen (16) pour extraire est réalisé de manière à lisser la fonction de détection dans le temps.

11. Dispositif selon la revendication 9 ou 10, dans lequel le moyen (16) pour extraire est réalisé de manière à trouver à un moment des maximums dans la fonction de détection (16c), et à utiliser un spectre de courte durée différentié ou un spectre de courte durée comme spectre significatif auquel est associé un moment auquel la fonction de détection présente un maximum.

12. Dispositif selon l'une des revendications 9 à 11, dans lequel le moyen (16) pour extraire est réalisé de manière à ne considérer comme significatifs que des maximums de la fonction de détection qui sont distants l'un de l'autre dans le temps de plus d'un laps de temps prédéfini.

13. Dispositif selon l'une des revendications 4 à 12, dans lequel le moyen (16) pour extraire est réalisé de manière à déterminer, comme succession de spectres de courte durée, des spectres de quantité et à utiliser des informations de phase des spectres de courte durée lors de l'extraction des spectres de courte durée significatifs.

14. Dispositif selon l'une des revendications précédentes, dans lequel le moyen (18) pour décomposer est réalisé de manière à additionner, pondérés, les spectres de courte durée extraits (18a), pour obtenir un nombre réduit de spectres de courte durée extraits.

15. Dispositif selon l'une des revendications 1 à 14, dans lequel le moyen (18) pour décomposer est réalisé de manière à effectuer, en vue d'une réduction de dimension, une analyse de composantes principales (18a), pour obtenir des spectres de courte durée traités.

16. Dispositif selon l'une des revendications précédentes, dans lequel le moyen (18) pour décomposer est réalisé de manière à effectuer une 'independent component analysis' (18b), pour générer une pluralité de signaux à composantes, à un signal à composantes étant associée une source d'information qui contribue au signal d'information.

17. Dispositif selon l'une des revendications précédentes, dans lequel le moyen (20) pour calculer la courbe enveloppante d'amplitude est réalisé de manière à multiplier une matrice comprenant les spectres de profil et une matrice comprenant une succession de spectres de courte durée du signal d'information, pour obtenir pour les sources sonores les courbes enveloppantes d'amplitude.

18. Dispositif selon l'une des revendications précédentes, dans lequel le moyen pour calculer la courbe enveloppante d'amplitude est réalisé, par ailleurs, de manière à déterminer une courbe enveloppante d'amplitude différenciée à l'aide des spectres de profil pour les sources sonores et à l'aide du spectrogramme différentiel.

19. Dispositif selon l'une des revendications précédentes, comprenant, par ailleurs, un moyen (22) pour classifier les signaux à composantes en signaux à composantes percussives et en signaux à composantes non percussives.

20. Dispositif selon la revendication 19, dans lequel le moyen (22) pour classifier est réalisé de manière à classifier sur base des spectres de profil et/ou des courbes enveloppantes d'amplitude.

21. Dispositif selon la revendication 19 ou 20, dans lequel le moyen (20) pour classifier est réalisé de manière à extraire une caractéristique des spectres de profil ou des courbes enveloppantes d'amplitude et à la comparer avec les caractéristiques de sources connues dans une banque de données.

22. Dispositif selon l'une des revendications précédentes, présentant, par ailleurs, un moyen (24) pour examiner les courbes enveloppantes d'amplitude pour une source sonore, pour accepter un maximum dans la courbe enveloppante d'amplitude comme insertion d'un signal de la source sonore lorsque le moyen (16) pour extraire avait extrait, à un moment similaire dans un seuil, un spectre de courte durée significatif.

23. Dispositif selon l'une des revendications précédentes, dans lequel le moyen (20) pour calculer la courbe enveloppante d'amplitude est réalisé de manière à calculer la courbe enveloppante d'amplitude pour une source sonore de sorte que la courbe enveloppante d'amplitude indique la manière dont varie dans le temps une intensité ou une pondération d'un spectre de profil de la source sonore.

24. Procédé pour analyser un signal d'information, aux étapes suivantes consistant à:

extraire (16) des spectres de courte durée significatifs ou des spectres de courte durée significatifs dérivés de spectres de courte durée du signal d'information d'une succession dans le temps de spectres de courte durée représentant le signal d'information, de la succession dans le temps étant extraits les spectres de courte durée qui se rapprochent plus d'une caractéristique spécifique que d'autres spectres de courte durée du signal d'information;
décomposer (18) les spectres de courte durée extraits en spectres de signal à composantes, un spectre de signal à composantes représentant un spectre de profil d'une source sonore qui génère un son correspondant à la caractéristique recherchée, et un autre spectre de signal à composantes représentant un spectre de profil d'une autre source sonore qui génère un son correspondant à la caractéristique recherchée; et
calculer (20) une courbe enveloppante d'amplitude pour les sources sonores, une courbe enveloppante d'amplitude pour une source sonore indiquant la manière dont varie dans le temps un spectre de profil de la source sonore, à l'aide des spectres de profil et d'une succession de spectres de courte durée représentant le signal d'information.

25. Programme d'ordinateur avec un code de programme adapté pour réaliser un procédé pour analyser un signal d'information selon la revendication 24 lorsque le programme d'ordinateur est exécuté sur un ordinateur.

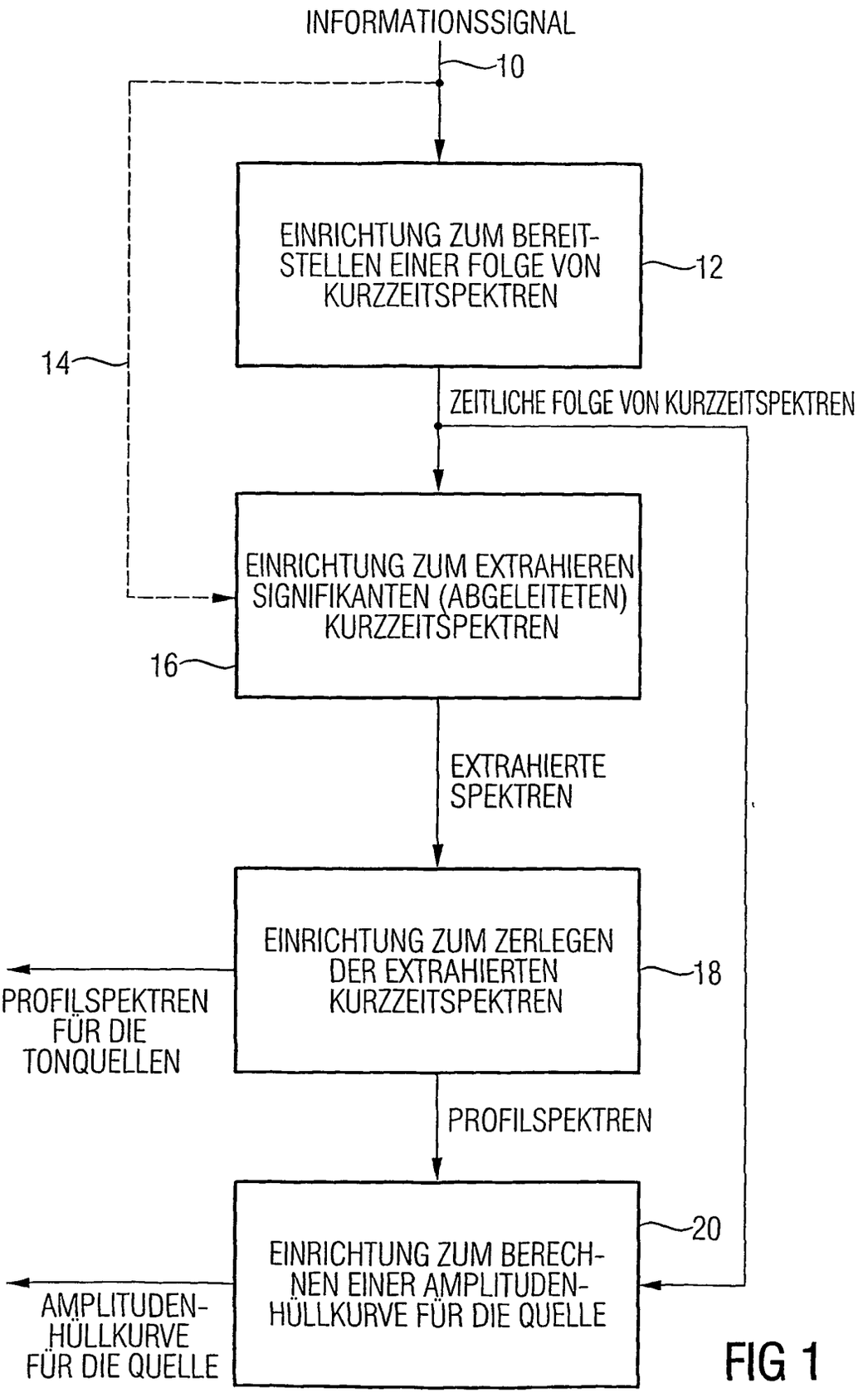
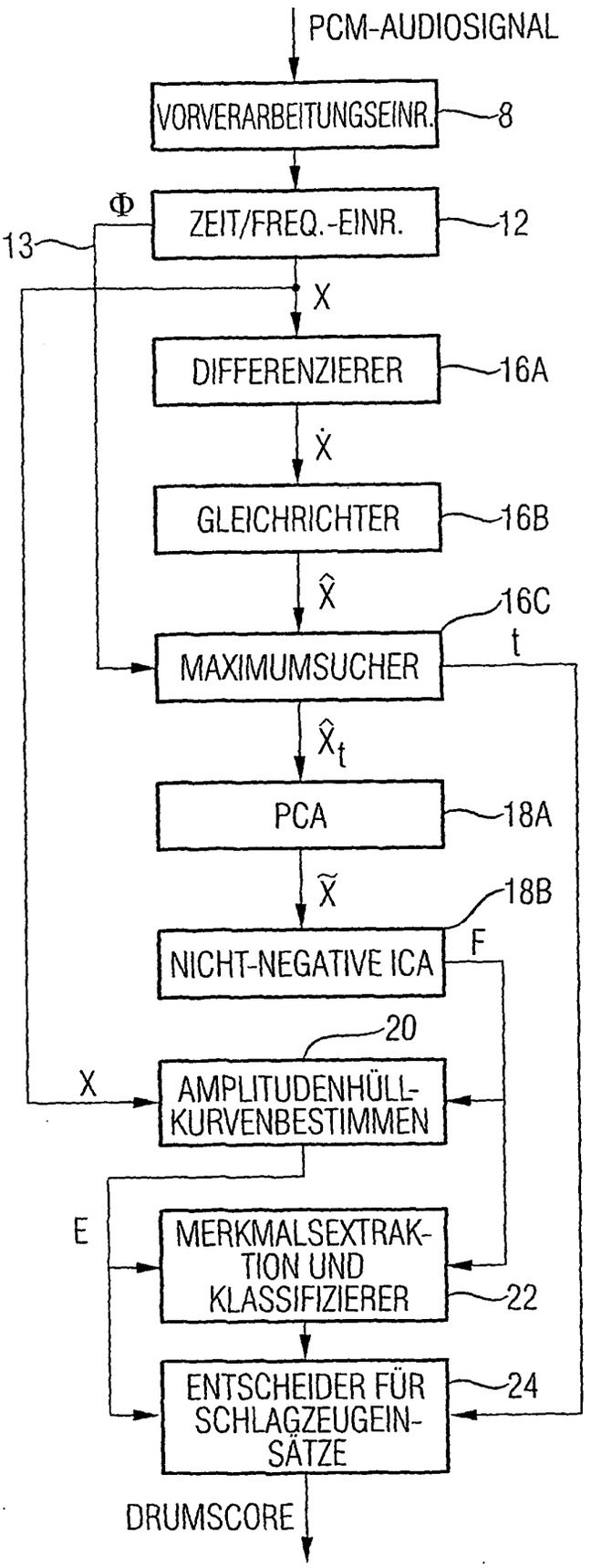

INFORMATIONSSIGNAL
10
EINRICHTUNG ZUM BEREIT-STELLEN EINER FOLGE VON KURZZEITSPEKTREN
12
14
ZEITLICHE FOLGE VON KURZZEITSPEKTREN
EINRICHTUNG ZUM EXTRAHIEREN SIGNIFIKANTEN (ABGELEITETEN) KURZZEITSPEKTREN
16
EXTRAHIERTE SPEKTREN
EINRICHTUNG ZUM ZERLEGEN DER EXTRAHIERTEN KURZZEITSPEKTREN
18
PROFILSPEKTREN FÜR DIE TONQUELLEN
PROFILSPEKTREN
EINRICHTUNG ZUM BERECH-NEN EINER AMPLITUDEN-HÜLLKURVE FÜR DIE QUELLE
20
AMPLITUDEN-HÜLLKURVE FÜR DIE QUELLE

FIG 1



F: PROFILSPEKTREN FÜR KOMPONENTEN

E: AMPLITUDENHÜLL-KURVEN FÜR KOMPONENTEN

FIG 2

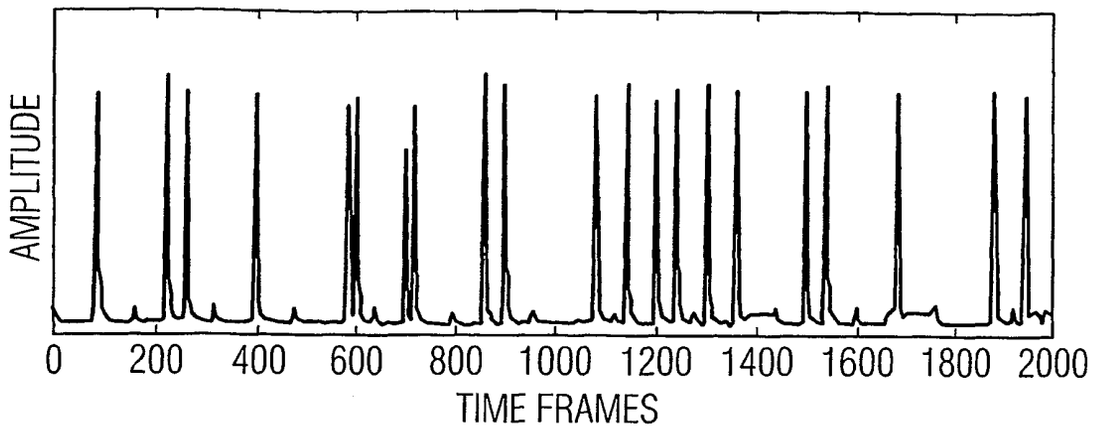


FIG 3A AMPLITUDENHÜLLKURVE FÜR EINE PERKURSSIVE QUELLE (KICK DRUM)

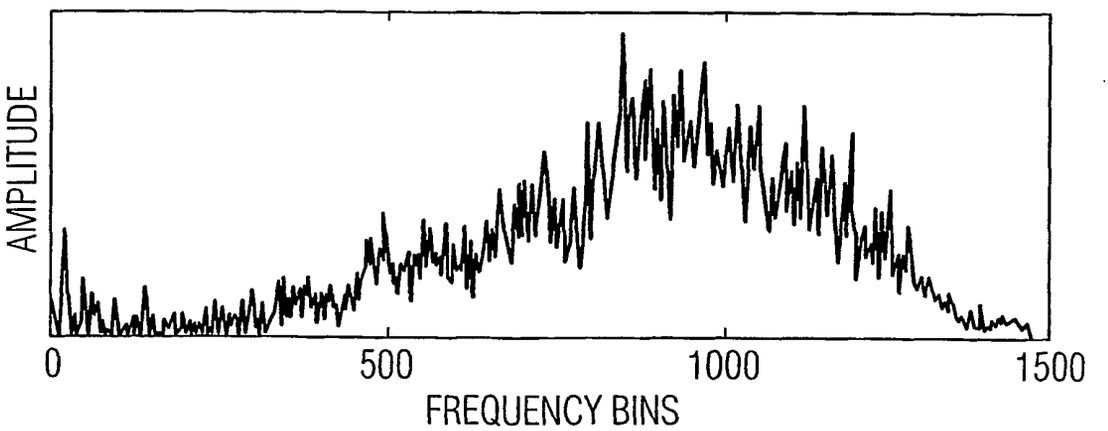


FIG 3B PROFILSPEKTRUM FÜR EINE PERKUSSIVE QUELLE (Hi-Hat)

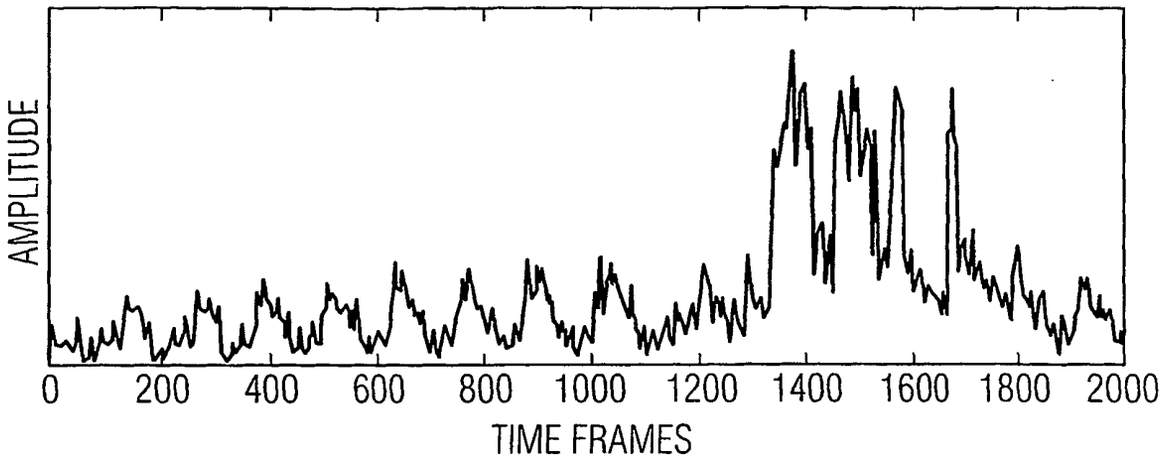


FIG 4A AMPLITUDENHÜLLKURVE FÜR EIN HARMONISCH AUSGEHALTENES INSTRUMENT (TROMPETE)

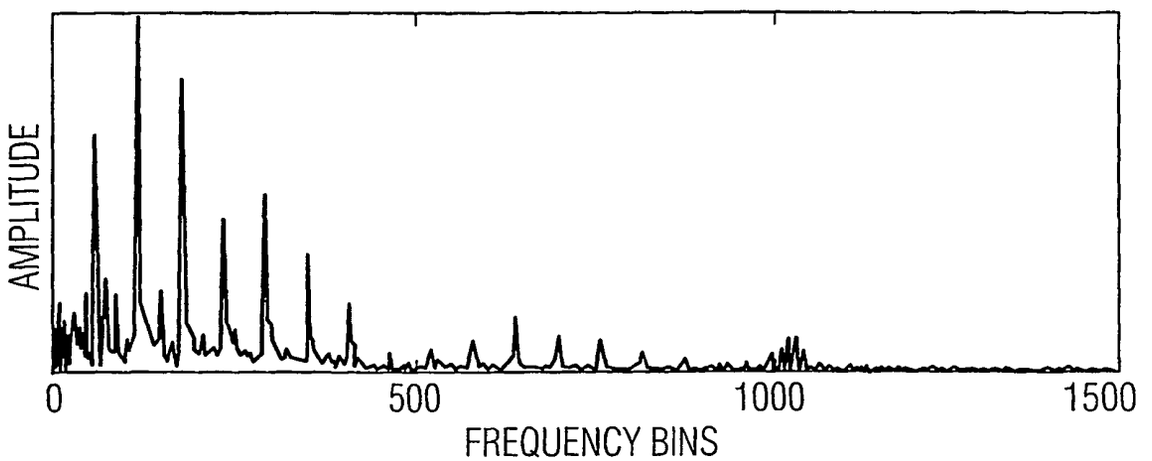


FIG 4B PROFILSPEKTRUM FÜR EIN HARMONISCH AUSGEHALTENES INSTRUMENT (GITARRE)

IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

Diese Liste der vom Anmelder aufgeführten Dokumente wurde ausschließlich zur Information des Lesers aufgenommen und ist nicht Bestandteil des europäischen Patentdokumentes. Sie wurde mit größter Sorgfalt zusammengestellt; das EPA übernimmt jedoch keinerlei Haftung für etwaige Fehler oder Auslassungen.

In der Beschreibung aufgeführte Nicht-Patentliteratur

- **M.A. CASEY ; A. WESTNER.** Separation of Mixed Audio Sources by Independent Subspace Analysis. *Proc. of the International Computer Music Conference,* 2000 **[0013]**
- **I.F.O. ORIFE.** Riddim: A rhythm analysis and decomposition tool based on independent subspace analysis. *Master thesis, Darthmouth College,* 2001 **[0013]**
- **C. UHLE ; C. DITTMAR ; T. SPORER.** Extraction of Drum Tracks from polyphonic Music using Independent Subspace Analysis. *Proc. of the Fourth International Symposium on Independent Component Analysis,* 2003 **[0013]**
- **D. FITZGERALD ; B. LAWLOR ; E. COYLE.** Prior Subspace Analysis for Drum Transcription. *Proc. Of the 114th AES Convention,* 2003 **[0013]**
- **D. FITZGERALD ; B. LAWLOR ; E. COYLE.** Drum Transcription in the presence of pitched instruments using Prior Subspace Analysis. *Proc. of the ISSC,* 2003 **[0013]**
- **M. PLUMBLEY.** Algorithms for Non-Negative Independent Component Analysis. *IEEE Transactions on Neural Networks,* Mai 2003, vol. 14 (3), 534-543 **[0013]**
- **T. HEITTOLA et al.** Locating Segments with Drums in Music Signals. *3RD INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL ISMIR,* 20. Oktober 2002, 1-6 **[0013]**