

(19)



(11)

EP 1 758 096 A1

(12)

EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:
28.02.2007 Patentblatt 2007/09

(51) Int Cl.:
G10L 11/00^(2006.01) G10H 1/00^(2006.01)

(21) Anmeldenummer: **05107730.3**

(22) Anmeldetag: **23.08.2005**

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

(72) Erfinder: **Schierle, Rainer**
9490 Vaduz (LI)

(74) Vertreter: **Büchel, Kaminski & Partner**
Patentanwälte Est.
Austrasse 79
9490 Vaduz (LI)

(71) Anmelder: **Schierle, Rainer**
9490 Vaduz (LI)

(54) **Verfahren und Vorrichtung zur Mustererkennung in akustischen Aufnahmen**

(57) Zur Mustererkennung in akustischen Aufnahmen wird ein aufgenommenes Signal in einzelne Frequenzbereiche zerlegt und nachfolgend zur spektralen Zerlegung in wenigstens eine Koeffizientendatei transformiert. Hierbei erfolgen parallel eine hinsichtlich der Frequenzauflösung optimierte erste Transformation und

eine hinsichtlich der Zeitauflösung optimierte zweite Transformation.

Auf der Basis der Koeffizientendatei erfolgt eine harmonische Dekomposition mit einer Musterzuordnung. Die identifizierten Muster können nachfolgend modifiziert und weiter genutzt werden, beispielsweise in Form einer graphischen Darstellung oder akustischen Wiedergabe.

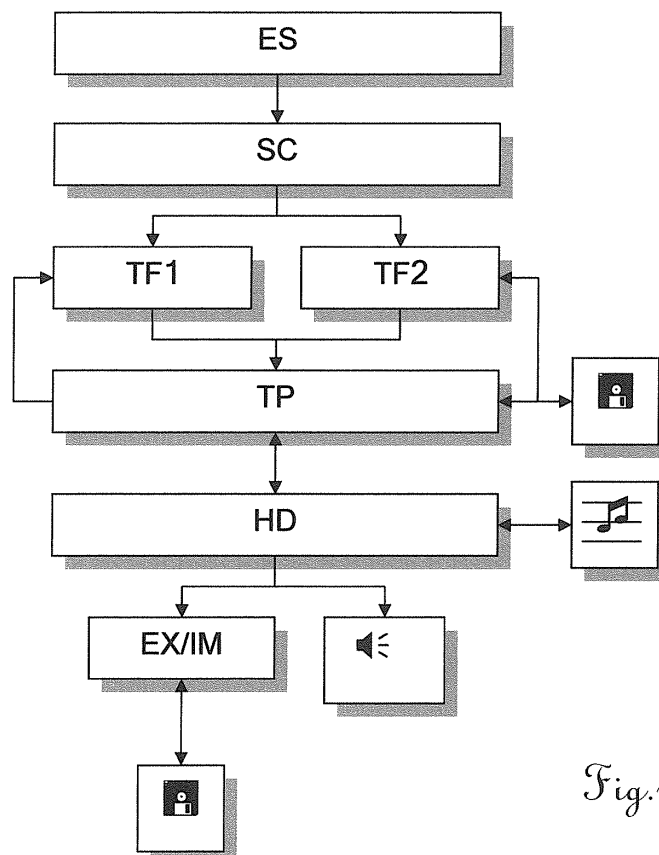


Fig.1

EP 1 758 096 A1

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur Mustererkennung in akustischen Aufnahmen nach dem Oberbegriff des Anspruchs 1 bzw. 13, sowie ein Computerprogrammprodukt und ein Datenstrukturprodukt.

[0002] In vielen Anwendungsbereichen besteht das Erfordernis, in Aufnahmen akustischer Signale Muster zu erkennen und zur Nutzung zu konvertieren. Beispiele hierfür stellen seismische Messungen, Schwingungsanalysen im Maschinenbau, die Selektion von Audiosignalen im Hörgerätebereich, die Sprachanalyse oder die Konversion von Musik in abspiel- bzw. veränderbare Formate dar. Die Grundproblematik in allen diesen Bereichen ist stets dieselbe, im folgenden wird rein exemplarisch die Mustererkennung in Aufnahmen von Musikstücken erläutert, ohne hierbei eine Einschränkung auf diesen Anwendungszweck zu begründen. Das erfindungsgemässe Verfahren bzw. die erfindungsgemässe Vorrichtung können auch zur Lösung anderer Problemstellungen, insbesondere aus den oben explizit dargestellten Gebieten, angewendet werden.

[0003] Zur Verarbeitung von akustischen Aufnahmen bzw. Audiosignalen werden diese heute in aller Regel digitalisiert. Beispielsweise erfolgt eine Aufnahme durch geeignete Sensoren, wobei das aufgenommene Signal abgetastet und digitalisiert gespeichert wird. Ein weit verbreiteter Ansatz ist die Konversion und Abspeicherung im WAVE-Format. Um eine für das menschliche Ohr verlustfreie Konversion und Speicherung zu ermöglichen, erfolgt zumeist ein Sampling 44,1 kHz und 16 Bit Auflösung, so dass für die vom menschlichen Ohr maximalen wahrnehmbaren Frequenzen das Nyquist-Theorem erfüllt ist.

[0004] In diesem Format sind somit zwar alle akustisch relevanten Anteile erfasst, so dass für das menschliche Ohr eine Wiedergabe ohne erkennbaren Verlust möglich ist. Jedoch erfordert dieses Format einen grossen Speicherplatz, was z.B. bei einer Übertragung im Internet nachteilig ist, da lange Übertragungszeiten die Folge sind. Zudem erfolgt keine Speicherung von aufgelösten Mustern, d.h. eine Trennung von z.B. verschiedenen Musikinstrumenten erfolgt nicht, so dass beispielsweise keine einfache Veränderung der Aufnahme möglich ist, z.B. durch Streichen eines Instrumentes.

[0005] Ein weiteres Datenformat, welches quasi den entgegengesetzten Informationsgehalt verkörpert, ist das MIDI-Format, wobei MIDI für Musical Instrument Digital Interface steht. Dieses für den Datenaustausch zwischen Synthesizern entwickelte Format überträgt statt Audiodaten Kontrollsignale, welche durch einen Synthesizer wiedergegeben oder auch graphisch oder visuell dargestellt werden können. Im weit verbreiteten GM-Standard erfolgt dabei eine Kodierung bzw. spätere Wiedergabe in 128 Klangfarben. Aufgrund des somit vergleichsweise geringen Datenumfangs eignet sich dieses Format gut zur Übertragung im Internet. Allerdings kann diese geringe Bandbreite an Klangfarben den natürlichen Klang nicht wiedergeben. Zudem besteht beim MIDI-Format eine Abhängigkeit der Wiedergabe von der Hardware.

[0006] Im Stand der Technik werden verschiedene Ansätze verfolgt, die eine Mustererkennung in Audiosignalen erlauben, wobei häufig eine Konversion von Wave- in MIDI-Dateien erfolgt.

[0007] Beispielsweise offenbart US 6,140,568 ein System und ein Verfahren zur automatischen Erkennung und Identifikation einer Vielzahl von Frequenzen, die gleichzeitig in einem Audiosignal enthalten sind, wie z.B. Zeitdauer, Amplitude und Phase dieser Frequenzen. Aus diesen Frequenzen werden zur Bestimmung der fundamentalen Frequenzen harmonische Komponenten herausgefiltert. Das System beinhaltet ein computerlesbares Medium mit ausführbarem Code zur Dekomposition des Signals in seine sinusförmigen Komponenten durch Berechnung und Vergleich zwischen dem Eingangssignal und sinusförmigen Wellen mit verschiedenen Kombinationen von Phase und Amplitude. Das System verwendet ebenfalls verschiedene Optimierungs- und Fehlerkorrekturroutinen.

[0008] In der Schrift US 6,355,869 B1 werden ein Verfahren und ein System zur Erzeugung von Noten aus einer Aufzeichnung von Musik sowie die Erzeugung eines editierbaren Musikformats beschrieben. Das Verfahren beruht auf dem Speichern der Musikaufnahme als Wave-Datei aus der für jeden relevanten Abschnitt in der Aufzeichnung eine Pseudo-Wave-Datei erzeugt wird. Für jede Pseudo-Wave-Datei wird eine Sequenz-Datei erzeugt, aus der wiederum eine Liste von Ereignissen generiert wird. Diese Liste wird in eine MIDI-Datei oder eine andere notenlesbare Datei konvertiert und zum Ausdruck der Noten in ein Noten-Programm importiert.

[0009] Während die Mustererkennung für verschiedene Typen von Mustern bzw. die Identifikation von vielen Musikinstrumenten mit den Ansätzen des Stands der Technik geleistet werden kann, bereiten einige Mustertypen nach wie vor Probleme. So können mit bisherigen Verfahren gerade die Schlagzeuganteile in Audiosignalen nur schlecht aufgelöst und in Noten dargestellt werden. Das Problem besteht beim Schlagzeug darin, dass dieses einen breiten Bereich an spektralen Beiträgen liefert, der mit den bisherigen Verfahren nicht eindeutig separiert und analysiert werden kann.

[0010] Ausserdem erlaubt datenhaltungsseitig das weitverbreitete MIDI-Format nur eine Speicherung bzw. Wiedergabe, welche starke Abstriche in Hinblick auf die originale Klangqualität mit sich bringt.

[0011] Die Aufgabe der vorliegenden Erfindung besteht daher in der Bereitstellung eines verbesserten Verfahren bzw. einer verbesserten Vorrichtung, welche auch die Auflösung von Komponenten mit einem breiten Bereich an spektralen Beiträgen ermöglicht.

[0012] Eine weitere Aufgabe der Erfindung besteht darin, auch eine Identifikation von Schlagzeuganteilen in Aufnahmen von Musik zu ermöglichen.

[0013] Eine weitere Aufgabe der Erfindung besteht darin, eine verbesserte interaktive Veränderbarkeit von akustischen Aufnahmen zu ermöglichen.

[0014] Eine weitere Aufgabe der Erfindung besteht in der Bereitstellung eines Datenstrukturproduktes, welches unter Abspeicherung von Kontrollsignalen eine möglichst originalgetreue Wiedergabe erlaubt, so dass beispielsweise die Vorteile von Wave- und MIDI-Format kombiniert werden, ohne deren Nachteile in Kauf nehmen zu müssen.

[0015] Diese Aufgaben werden erfindungsgemäss durch die Merkmale der Ansprüche 1 und 13 bzw. durch die kennzeichnenden Merkmale der abhängigen Ansprüche gelöst oder die Lösungen fortgebildet.

[0016] Das erfindungsgemässe Verfahren bzw. die erfindungsgemässe Vorrichtung zur Mustererkennung in akustischen Aufnahmen analysieren akustische Signale, wie sie beispielsweise durch Mikrophone erfasst werden. Diese Signale können Musikstücke, Sprache, Maschinenvibrationen, seismische Schwingungen oder andere Formen von mechanischen Schwingungen darstellen.

[0017] Das Signal wird nach oder während der Aufnahme vorzugsweise digitalisiert, um eine Signalverarbeitung auf Rechnern zu erlauben, wobei die Datenspeicherung z.B. im Wave-Format erfolgen kann. Alternativ oder ergänzend ist eine Realisierung des Verfahrens auch in Analogtechnik möglich, z.B. durch eine entsprechende Schaltung.

[0018] Das erfasste und gespeicherte Signal wird nachfolgend in einzelne Frequenzbereiche, z.B. Oktaven, zerlegt, wofür an sich bekannte Verfahren zur Anwendung kommen können. Ein Beispiel hierfür ist die Pyramidendekomposition, bei der das Eingangssignal in verschiedene Subband-Signale mit unterschiedlichen Frequenzbereichen zerlegt wird. Typischerweise umfasst das erste Subband nur die höchsten Frequenzen. Die nachfolgenden Subbänder beinhalten dann die jeweils nächsten niedrigeren Signalanteile.

[0019] Die Frequenzbereiche werden nachfolgend spektral zerlegt, woraus jeweils ein Satz von Koeffizienten folgt. Erfindungsgemäss erfolgt diese spektrale Zerlegung in zwei voneinander unabhängigen Transformationsprozessen.

[0020] Hierfür geeignete Transformationsalgorithmen stellen beispielsweise die Fourier-Transformation, Fast-Fourier-Transformation, Wavelet-Transformation, SinusTransformation oder Kosinus-Transformation dar, wobei insbesondere die diskreten Varianten geeignet sind.

[0021] Einer der beiden voneinander unabhängigen Transformationsprozesse ist hinsichtlich der zeitlichen Auflösung optimiert. Hierfür wird das zeitliche Fenster vergleichsweise kurz gewählt, so dass der zeitliche Verlauf gut aufgelöst wird. Die zeitliche Beschränkung vermindert jedoch die Frequenzauflösung, so dass der andere Transformationsprozess den gleichen Frequenzbereich mit einem vergleichsweise grossen zeitlichen Fenster analysiert, so dass hierfür eine höhere Auflösung der Frequenzen erfolgt. Beide Transformationen liefern jeweils einen Koeffizientensatz für die beitragenden Frequenzanteile. Das entstandene TF-Ausgangsbild (TF für Frequency-Time-Image) wird nun seinerseits in Subbänder über Zeit und/oder Zeit und Frequenz zerlegt, was wiederum einer Transformation mit längeren Zeitkonstanten entspricht. Verschiedene Frequenz-Zeit-Bilder (TF) werden dazu verwendet, um Signale oder Signaleigenschaften zu detektieren und um Originalsignale (Eingangssignale) zu rekonstruieren.

[0022] Diese Transformationen sind somit für verschiedene Aufgabenbereiche optimiert, wie z.B. die Unterteilung in perkussive und harmonische Signalkomponenten. Als mögliche Transformation sei rein exemplarisch die Fourier-Transformation beschrieben:

$$A_s(t, f) = \int I(t) \sin(\omega t) dt \quad (1)$$

$$A_c(t, f) = \int I(t) \cos(\omega t) dt \quad (2)$$

wobei

$A_s(t, f)$ den Sinusanteil des Ausgangssignals,
 $A_c(t, f)$ den Cosinusanteil des Ausgangssignals,
 ω die Kreisfrequenz der zu untersuchenden Frequenzkomponente und
 t die Zeit bezeichnen.

[0023] Das nach der Transformation in den Schichten der Ausgangsmenge gespeicherte Signal ist eine Mischung aus den Transformationsausgangssignalen und einer Pyramidenzerlegung der jeweils nächsthöheren Stufe der Pyramide

$$TF(n, t, f) = A_s(t, f) \otimes A_c(t, f) \quad (3)$$

wobei $\otimes$ eine allgemeine Verknüpfung bezeichnet, die im einfachsten Falle einer Addition entspricht. Werden auch Beiträge nächsthöherer bzw. darüberliegender Schichten berücksichtigt so ergibt sich

$$TF(n, t, f) = A_s(t, f) \otimes A_c(t, f) \otimes TF(n-1, t, f) \quad (3a)$$

wobei $TF(n-1, t, f)$ den Beitrag der nächsthöheren Schicht $n-1$ bezeichnet. Auch können $A_s(t, f)$ und $A_c(t, f)$ im üblichen Fall die Amplituden und Phasenwerte der Fouriertransformation darstellen

$$Amp(t, f) = \sqrt{A_s(t, f)^2 + A_c(t, f)^2} \quad (4)$$

bzw.

$$\varphi = \text{atan}\left(\frac{A_s(t, f)}{A_c(t, f)}\right) \quad (5)$$

[0024] Die einzelnen Schichten der Pyramide können aus einer Kombination von Hoch-, Tiefpassfiltern und Subsampling erzeugt werden. Diese TF-Pyramide kann auch mehrfach vorhanden generiert werden, um verschiedenen Zwecken, wie Signalanalyse und Signalrekonstruktion Rechnung zu tragen.

[0025] Informationen aus einer oder mehrerer der Schichten werden in einem Filter, beispielsweise einem zweidimensionalen Filter mit Mittelwertkern, zu einem eindimensionalen Vektor zusammengefasst, aus welchem dann beispielsweise mit Detektion lokaler Maxima Notenereignisse ableitbar sind.

[0026] Zusätzlich zur Anordnung mit zwei Transformationen, welche beispielsweise für harmonische und perkussive Signale optimiert sind, kann auch ein Schema eingesetzt werden, bei dem eine oder mehrere Transformationen einen mehrschichtigen Ausgangsbereich füllen. Dies bedeutet, dass für jede Oktave des Subband-Eingangssignals eine Transformation für eine (1) bis mehrere (12 für eine Oktave mit Halbtönen, 14 oder 16, um mit Filtern in Frequenzrichtung filtern zu können) Frequenzen durchgeführt wird, was ein Frequenz/Zeit-Bild erzeugt. Dieses Bild kann aus dem Signal einer oder mehreren Transformationen erstellt werden. So können beispielsweise Anteile aus der frequenzoptimierten Transformation mit Anteilen der perkussiven Transformation so gemischt werden, dass eine klare Abgrenzung zwischen harmonischen und perkussiven Signalen möglich wird.

[0027] Nach dem Transformieren der Frequenzbereiche wird die spektrale Zerlegung durch Erzeugung wenigstens einer Koeffizientendatei abgeschlossen. In diese Koeffizientendatei werden Koeffizienten aus den Koeffizientensätzen der beiden Transformationen übernommen, wobei die Koeffizienten aus einer der beiden Sätze ausgewählt oder aber auch als Mischung von Koeffizienten erzeugt werden können. Somit werden die beiden Koeffizientensätze der unterschiedlichen Transformationen in einer Gesamttransformation unter Selektion oder Mischung in eine Koeffizientendatei überführt, wobei diese Datei dann Anteile aus beiden Transformationen enthält.

[0028] Die Erzeugung der Koeffizientendatei nutzt dabei Heuristiken, vorgegebene Informationen, z.B. aus früheren Analysen, oder auch statistische Auswertungen des aktuellen Signals. Grundsätzlich werden alle Frequenzbänder durch beide Transformationsprozesse geführt. Jedoch kann auch, z.B. aufgrund von vorgegebenen Informationen, für einzelne Frequenzbänder nur einer der beiden Transformationsprozesse zur Anwendung kommen, so dass nur das Ergebnis dieses Schrittes weiterverwendet wird.

[0029] Die Auswahl bzw. Mischung von Koeffizienten zur Erzeugung kann mittels verschiedener Verfahren erfolgen.

[0030] In einem Ansatz erfolgen eine erste Fouriertransformation mit langem zeitlichem Fenster und eine zweite Fouriertransformation mit kurzem zeitlichen Fenster und nachfolgendem Tiefpassfilter. Für die Ergebnisse beider Transformationen wird jeweils der Realteil berechnet und deren Verhältnis gebildet. Anhand dieses Verhältnisses wird entschieden, aus welcher Transformation der Koeffizient gewählt wird.

[0031] Ein anderer Ansatz beruht auf der Analyse der Steigung in einer Darstellung von Phase gegenüber Frequenz, d.h. der frequenzabhängigen Steigung des Phasensignals. Durch das Setzen von Schwellen oder das Berechnen eines Gewichtungsparmeters erfolgt eine Bestimmung, welcher Koeffizient verwendet wird, bzw. ob und wie eine Mischung von Koeffizienten erfolgt.

[0032] Die Nutzung von vorgegebenen Informationen erfolgt durch einen Vergleich der durch die Transformationen erhaltenen Koeffizientensätze mit einem Satz gespeicherter Koeffizienten. Dieser Vergleich dient als Auswahlkriterium für die Koeffizienten bzw. deren Mischung.

[0033] Durch den vollständigen Transformationsprozess wird schliesslich eine Datei erzeugt, welche die ausgewählten bzw. gemischten Koeffizienten enthält. Darüber hinaus kann in dieser Datei noch statistische Information bezüglich des Signals abgelegt sein.

[0034] Auf der Basis dieser Koeffizientendatei erfolgt die harmonische Dekomposition, die schliesslich zu einer Zuordnung von spektralen Anteilen zu Mustern, wie z.B. speziellen Musikinstrumenten, führt. Die detektierten Muster oder Ereignisse können nach einer Konversion graphisch dargestellt, z.B. als Noten, oder durch Synthesizer wiedergegeben werden. Unter Mustern oder Ereignissen sollen hier die charakteristischen Anteile in einem akustischen Signal verstanden werden, deren Identifikation das Ziel der Analyse darstellt. Dies können beispielsweise einzelne Musikinstrumente, Worte oder seismische Kenngrössen sein.

[0035] Grundlage der Dekomposition bilden erfindungsgemäss nicht nur die Koeffizienten selbst, sondern auch deren Aggregate, z.B. das zeitliche Integral einer Amplitude für eine bestimmte Frequenz, oder statistische Informationen.

[0036] Zur Dekomposition kann im einfachsten Fall ein Vergleich mit einer Datenbank erfolgen, in der Beispiele für Muster abgelegt sind. Solche Datenbanken stehen beispielsweise für Musikinstrumente zur Verfügung.

[0037] Eine weitere Möglichkeit stellt der Aufbau eines Modells für die zu identifizierenden Muster dar, wobei dieses Modell z.B. aus dem aktuellen Signal mit statistischen Verfahren aufgebaut werden kann. Das Modell wird iterativ mit dem Signal verglichen und schrittweise optimiert. Unterschreitet das verbleibende Residuum einen vorgegebenen Schwellwert, wird das Verfahren abgebrochen und die Mustererkennung wird als hinreichend gut betrachtet.

[0038] Zur Merkmals- bzw. Notenerkennung können verschiedene Ansätze alternativ oder kumulativ Verwendung finden.

[0039] So werden beispielsweise durch geeignete ein- oder zweidimensionale Filter in den einzelnen Schichten der TF-Pyramide charakteristische Merkmale der einzelnen Musikinstrumente ermittelt. Diese Merkmale können dann direkt den einzelnen Musikinstrumenten und deren Repräsentation im Notationsformat (z.B. Midi, oder internes Format) zugewiesen werden. Alternativ werden die Merkmale einem neuronalen Netzwerk als Eingangsvariablen zugeführt.

[0040] In diesem neuronalen Netzwerk werden die durch die Merkmale bestimmten Regionen der TF-Pyramiden genauer untersucht, beispielsweise durch Pixel-zu-Pixel-Vergleich in einer abgegrenzten Umgebung des Merkmals. Die ermittelten Resultate dieser Vergleiche können rückgekoppelt auf die Merkmalerkennung eine Verbesserung der Merkmalerkennung bewirken. Beispielsweise werden Merkmalszentren, Merkmalsschwellwerte und Frequenz-Zeit-Ausdehnung der Merkmalerkennung angepasst. Mit diesen Methoden lassen sich Merkmale für perkussive und/oder harmonische Klänge bestimmen. Im speziellen werden dadurch einzelne Töne eines Instruments, z.B. Gitarre, Bass, Trommeln und Becken eines Schlagzeugs, aber auch Piano- und Gitarrenakkorde erkannt. In grundsätzlich gleicher Weise können auch seismische Ereignisse oder sprachliche Merkmale, z.B. auszublenkende Hintergrundgeräusche bei einer akustischen Kommunikationsverbindung, analysiert werden.

[0041] Da sich in den Eingangssignalen Merkmale oft wiederholen, können ermittelte Merkmale und Muster dazu verwendet werden, den gesamten Informationsgehalt (TF) auf solche Wiederholungen zu durchsuchen.

[0042] Die ermittelten Muster werden nach vorgegebenen Kriterien oder nach der Analyse durch Zuordnung klassifiziert, wobei diese Zuordnung vom Computerprogramm vollautomatisch, halbautomatisch bzw. interaktiv durch den Programmanwender durchgeführt werden kann. Zur Verbesserung der Klassifizierung der Muster kann die Ergebnismenge (TF) nochmals auf vergleichbare Muster untersucht werden. Diese Methode ist zeitsparend, da die Transformation oftmals ein vergleichsweise langer andauernder Prozess sein kann.

[0043] Alle Verfahren des Stands der Technik zur Musikererkennung führen bislang zu einem statischen, nicht interaktiv korrigierbaren Notenbild, welches fehlerbehaftet oder nicht korrekt im Sinne der gewünschten Darstellung ist. Zur Verbesserung stehen erfindungsgemäss Methoden zur Verfügung, welche durch interaktive Vorgabe von Parametern zwischen dem Computerprogramm und dem Anwender die generierte Notendarstellung modifizierbar gestalten. Beispielsweise können durch Informationen mit zeitlichem Charakter identifizierte Harmonien (z.B. Gitarren- und Pianoakkorde) verbessert oder geändert werden.

[0044] So kann beispielsweise die Takteinteilung als zeitliche Klassifizierung manuell ergänzt oder geändert werden. Notenschrift bedarf einer Klassifizierung in zeitlichem Sinne in einer solchen Weise, dass ermittelten Notenwerten Notentlängen zugeordnet werden können. Eine Funktion im Anwenderprogramm ermöglicht hierbei die Markierung von Taktbeginn und eine automatische Funktion des Programms ermittelt dann zwischen diesen Markierungen die fehlenden Takte. Dieser Prozess kann wiederholt werden, bis die Takteinteilung zufriedenstellend ist. Es können aber auch Funktionen verwendet werden, welche die Takteinteilung automatisch erkennen.

[0045] Eine Verbesserung der Harmonieerkennung durch zeitliche Klassifizierung ist aufgrund einer Einteilung des Informationsgehalts in Takte möglich, die zur Verbesserung der Harmonieerkennung herangezogen werden kann, indem von der Tatsache Gebrauch gemacht wird, dass sich in real gespielter Musik die Harmonien oft beim Taktwechsel ändern.

[0046] Eine unzureichende Einstellung von automatischen oder manuellen Schwellwerten bei der Notenerkennung des Stands der Technik führt dazu, dass der zeitlich aufwendige Prozess der Notenerkennung neu gestartet werden muss. Erfindungsgemäss können Schwellwerte für die Notenerkennung auch nachträglich verändert werden, damit die erkannten Noten dem Anwender in optimaler Darstellung zur Verfügung gestellt werden können. Dazu werden Kriterien, beispielsweise Merkmale, mit einem Schwellwert so versehen, dass Signale unter den Schwellwert nicht als Musiknoten dargestellt werden und auch nicht erklingen.

[0047] Dabei kann der Nutzer durch Interaktion mit dem System auch rückgekoppelt auf das Ergebnis einwirken. Beispielsweise kann dieser aus seiner - z.B. durch das Anhören des aufgenommenen Musikstückes erhaltenen - Kenntnis der Besetzung einer Musikgruppe eine Vorauswahl der vorhandenen Musikinstrumente manuell vorgeben. Durch diese vorgegebenen Informationen wird dann die harmonische Dekomposition bzw. die Mustererkennung erleichtert und beschleunigt. Die Basis dieser Modifizierbarkeit stellt somit das erfindungsgemässe Verfahren dar, dass eine Modellbildung mit veränderbaren Koeffizienten beinhaltet, welche im Stand der Technik nicht geleistet wird bzw. werden kann.

[0048] Um eine optimale Nutzung und interaktive Veränderbarkeit zu gewährleisten erfolgt eine angepasste Darstellung der Ergebnisse mit verschiedenen Elementen. Zur Auswahl und Veränderung von Ereignissen wird ein Ereignisbild, beispielsweise als Bild mit notationsüblichen, in Y-Richtung angeordneten Gruppen von Linien, welche Tonhöhen entsprechen, generiert. In X-Richtung wird die Zeit aufgetragen oder eine zur Zeit proportionale Grösse. Ereignisse werden durch Notenköpfe oder aber ganz allgemein durch Symbole eines Fonts oder aber Bitmap oder anderen graphischen Formaten erhältliche Muster oder Bilder angezeigt. Dabei wird die Y-Position im Bild durch die Zuordnungstabelle oder einer mathematischen Funktion der Eigenschaften des Ereignisses zugeordnet, z.B. die Notenhöhe D6 (Midi 74) als zweite Linie von oben).

[0049] Sobald die Takte festgelegt sind, können die Ereignisse auch in üblicher Musiknotenschrift dargestellt werden.

[0050] Eine Darstellung kann auch in Form von Leadsheets als einbis mehrseitigen Zusammenfassungen eines Musikstücks erfolgen. Leadsheets in traditionellem Sinne werden von Hand erzeugt. Mit dem erfindungsgemässen Verfahren kann nun auch ein automatisches Erzeugen von Leadsheets durchgeführt werden. Dazu werden im Musikstück Markierungen gesetzt, welche abgrenzbare Bereiche des Musikstücks beschreiben, z.B. Einleitung, 1. Strophe, 1. Refrain, Zwischenteil, etc. Das Verfahren erzeugt dann aus den ermittelten Noten, Takten, und Akkorden eine zusammengefasste Darstellung des gesamten oder eines Teils des Musikstücks. Dieser Darstellung kann dann noch der Liedtext angefügt werden, wobei dieser dann auch im Notenbild zusätzlich einfügbar ist.

[0051] Durch einen Schwellwertregler für Tonhöhe können Notenwerte aktiviert, zur Darstellung und zum Erklingen gebracht werden. Dabei kann festgelegt werden, ob Ereignisse ausgeblendet werden oder aber auch die Tonhöhe um einen bestimmten Betrag, beispielsweise eine Oktave, verschoben werden sollen, wodurch die Noten dann eine Oktave tiefer abgespielt und notiert werden. Hierdurch kann das Ergebnis in soweit verbessert werden, dass, wenn Noten durch ihre harmonischen Anteile erkannt werden, diese auf die Grundfrequenz transponiert werden können.

[0052] Mit geeigneten Auswahlinstrumenten, wie z.B. einer Maus, einem Keyboard oder einem anderen Tool, können einzelne oder Gruppen von Noten selektiert und ggf. nachfolgend, z.B. per Midi, abgespielt werden. Erfindungsgemäss besteht die Möglichkeit, die Originalklänge, welche zur Erstehung des Ereignisses geführt haben, zu rekonstruieren und über das Musiksystem des Computers wieder abzuspielen. Diese Rekonstruktionen können nun auch separat in Musikdateien abgelegt werden.

[0053] Zum weiteren Trennen in verschiedene Musikinstrumente können mit den genannten Methoden Notenereignisse selektiert und auf andere Tonspuren kopiert oder verschoben werden.

[0054] Zur Verbesserung des Schlagzeugergebnisses als sich wiederholende Abfolge mit Akzentuierung stehen Verfahren zur Verfügung, welche eine Korrelation sich wiederholender Muster feststellen können, wobei die Korrelationslänge automatisch durch die Algorithmen des Programms oder durch den Anwender oder durch die Festlegung der Takte ermittelbar ist. Durch diese Korrelation können auch verschiedene Teile eines Musikstücks identifiziert werden. Die so ermittelten Schlagzeugmuster werden zusammengefasst auch auf den Leadsheets notiert.

[0055] Mit der zuvor angeführten Methode der Schlagzeugnotenerkennung können Bereiche in TF-Schichten markiert werden, aus deren Umgebung Muster abgeleitet werden können. Ein Teil oder alle diese Muster werden miteinander verglichen, wobei beispielsweise die Methode der Summe der Quadrate der Differenzen übereinander gelegter Pixel als Kriterium herangezogen werden kann, was für den statischen Fall wie folgt formuliert werden kann

$$S = \sum_{t_1}^{t_2} \sum_{f=0}^{f_{\max}} (P(t, f) \otimes R(t, f))^2 \quad (6)$$

wobei der korrespondierende dynamische Fall gemäss

$$S(t_0) = \sum_{t_1=t_0}^{t_2-t_0} \sum_{f=0}^{f_{\max}} (P(t-t_0, f) \otimes R(t, f))^2 \quad (6a)$$

formuliert werden kann. Hierbei bezeichnen P ein Signalmuster und R ein Referenzmuster. Als Verknüpfungen $\otimes$ können beispielsweise Subtraktion oder Multiplikation verwendet werden. Das Referenzmuster kann ein Muster an einer anderen Stelle der TF-Matrix sein oder ein vorabgespeichertes Muster oder aber ein Muster, welches aus einer Kombination bestehender Muster, beispielsweise durch Mittelwertbildung, entstanden ist. Im dynamischen Fall werden beide Muster gegeneinander zeitlich verschoben, so dass eine zeitabhängige Übereinstimmung ableitbar ist. Bei kleinen Werten von S besteht eine grosse Ähnlichkeit der zu vergleichenden Muster. In einer aus Vergleichen aller Muster miteinander erstellten Matrix AS sind die Elemente $AS(i, j) = S(i, j)$.

[0056] Zur Klassifizierung werden Gruppen gebildet und einem Graphen zugeordnet. Hierbei besteht eine Verbindung von jedem Muster zu dem Muster, welches am ähnlichsten ist. Auf Grund von vorprogrammierten Merkmalen werden die Muster dann klassifiziert und Notenwerte zugeordnet.

[0057] Die Erkennung von Akkorden in Musikstücken erfolgt auf die gleiche Art wie oben beschrieben für Schlagzeugnoten mit Mustererkennung.

[0058] Die Erkennung von harmonischen Klängen, wie z.B. Gitarre, Bass, Piano, Melodie oder Gesang, nutzt Schwellwerte. Ein Schwellwert bestimmt dabei, ob eine Frequenz einer TF-Schicht aktiv ist oder nicht. Im einfachsten Fall wird jede aktive Frequenz in eine Note umgewandelt, wobei Position, Notenhöhe und Länge, d.h. der Eintritt über die Schwelle bis zum Austritt beim Übergang von aktiv zu unterhalb der Schwelle, bestimmt werden. Diese Methode wird beispielsweise zur Erkennung von Instrument herangezogen, welche nur wenige Obertöne erzeugen, wie z.B. eine Sinusorgel.

[0059] Für harmonische Signale mit hohem Obertonanteilen, d.h. die Töne liegen bei Frequenzanteilen, die ein Vielfaches der Grundfrequenz betragen werden für eine oder mehrere Schichten der TF-Pyramide die Produkte

$$F_0 \rightarrow F_0 \otimes (H_1 + H_2 + H_3 + \dots H_n) \quad (7)$$

mit F_0 als Grundfrequenz und $H_1, H_2, H_3, \dots H_n$ als Höherharmonischen, d.h. $H_1 = 2 \cdot F_0$, $H_2 = 3 \cdot F_0$ etc., gebildet, wobei als Verknüpfung $\otimes$ beispielsweise eine Multiplikation gewählt werden kann. Danach werden die Bereiche aktiviert, die einen zuvor ermittelten oder festgelegten Schwellwert überschreiten, als Ereignisse ermittelt und in Noten umgewandelt.

[0060] Zudem können Notenobjekte gesammelt werden. Jeder Note stehen typischerweise folgende Eigenschaften zu:

- o Position im Lied
- o Länge des Ereignisses
- o Text
- o Frequenz
- o Notenhöhe
- o Detektionsvolumen
- o Musikinstrument
- o Amplitude
- o Koeffizienten.

[0061] Hierfür können Sammlungen (Collections) von Noten angelegt werden, welche typischerweise nach Instrumenten in Tonspuren aufgeteilt werden. Diese Sammlungen können in Dateien auf einem Computersystem abgespeichert werden. Solche Dateien können auch über das Internet, drahtgebunden oder durch elektromagnetische Übertragung weitergereicht werden. Als Beispiele für Übertragungsprotokolle seien Http, Tcp, Https, SOAP, etc. angeführt, wobei aber auch andere Formate möglich sind.

[0062] Die ermittelten Ereignisse bzw. Noten werden auf eine oder mehrere Arten angezeigt. Beispielsweise stellt ein Ausführungsbeispiel die Ereignisse als eine Kombination von Symbolen (Notenköpfen) dar, wobei die vertikale Achse einen üblichen Notenbild und die horizontale Achse der Zeit entspricht. Da bei einem standardgemässen Notenbild mit 5 Zeilen jede Zeile für 3 Noten stehen kann (z.B. g, ges und gis) können diese Zustände durch verschiedene Symbole dargestellt werden, z.B. ein regulärer Notenkopf für g, ein Dreieck mit Spitze nach unten für ges und ein Dreieck mit

Spitze nach oben für gis. Zusätzlich kann die Ereignislänge durch ein Rechteck angezeigt werden. Eine weitere mögliche Darstellung der Ergebnisse ist die gebräuchliche Notenschrift.

[0063] Im Gegensatz zum erfindungsgemässen Verfahren, das eine Anpassung der Ergebnisse erlaubt, haben Verfahren des Stands der Technik den Nachteil, dass Schwellwerte vor der zeitaufwendigen Analyse gesetzt werden müssen. Bei unzureichender Einstellung muss der gesamte Analysevorgang wiederholt werden, was aufwendig, wenig benutzerfreundlich, fehleranfällig und zeitraubend ist. Das erfindungsgemässe Verfahren hat den Vorteil, dass Schwellwerte für die Notenerkennung auch nach der Analyse gesetzt werden können. Dadurch können die Ergebnisse in Echtzeit an die Wünsche des Benutzers angepasst werden. Dieses Verfahren kombiniert die Möglichkeiten der Notenerkennung mit der Notendarstellung in einer Weise, die es erlaubt, die Ergebnisse durch Interaktion des Programm Benutzers mit der Analysesoftware individuell anzupassen.

[0064] Mit der speziellen Anwendermethode des halbautomatischen Setzens der Taktstriche können Positionen im Ereignisbild markiert werden, die musikalisch den ersten Schlag eines Taktes markieren. Bei diesem Ansatz wird mindestens ein Takt durch zwei Markierungen gesetzt und so eine zeitliche Information vorgegeben. Das Programm errechnet dann automatisch, z.B. mit Hilfe von Extrapolation, die fehlenden Takte für das ganze Lied. Dabei entstehen durch die Ungenauigkeit des gesetzten Taktes und durch Tempovariationen im Lied oft Abweichungen vom Idealergebnis, d.h. der Annahme, dass alle Takte richtig gesetzt sind. Zusätzliche erste Schläge eines Taktes können vom Anwender gesetzt werden, wobei dann das neue Taktlayout jeweils neu berechnet wird.

[0065] Der oben dargestellte Schwellwertregler kann auch als Tonhöhenfilter eingesetzt werden, d.h. als Instrument, um Grenzfrequenzen festzulegen, wobei dann Notenereignisse mit Tonhöhen über (bzw. unter oder zentriert um) einem Schwellwert nicht angezeigt oder eben angezeigt und gespielt werden. Wahlweise können Noten, die ausserhalb der Schwelle liegen, durch Tonhöhentransposition (Oktaveverschiebung) wieder in den Bereich der angezeigten Ereignisse gebracht werden. Als Beispiel gelte ein Tiefpass, bei dem Noten über dem Wert 60 (mittleres C (C5) gemäss Midi Standard, 61 = cis5) nicht angezeigt werden. In einem Fall wird eine Note der Tonhöhe 70 nicht mehr angezeigt und/oder gespielt, im anderen Fall wird die Note um eine Oktave nach unten transponiert (70-12 Halbtöne = 58), somit wird die Note mit Tonhöhe 58 gezeigt und gespielt. Dieses Verfahren dient zur Verminderung von fälschlich erkannten Oktavesprüngen in Melodien, in denen die harmonischen Signale anstatt der Grundtöne erkannt wurden.

[0066] Im Rahmen der Transformation oder der harmonischen Dekomposition können darüber hinaus noch weitere Verfahren zum Einsatz kommen. So können beispielsweise die Koeffizienten benachbarter Frequenzen durch Interpolation oder durch statistische Verfahren erhalten werden.

[0067] Gleichfalls können Koeffizienten ergänzt oder ersetzt werden, indem synthetisch erzeugte Koeffizienten sowie solche aus früheren Aufnahmen, einer früheren Analyse des gleichen Signals oder Mischungen derselben verwendet werden. So können z.B. für eine Trommel obere Frequenzanteile künstlich aus einer Datenbank ergänzt werden.

[0068] Die erzeugten Koeffizientendateien können in einem eigenen Format oder aber - gegebenenfalls nach einer Konversion - auch in einem verbreiteten Datenformat, wie z.B. MIDI- oder Wave-Format exportiert werden. Gleichermassen können auch solche Dateien importiert und deren Inhalt im erfindungsgemässen Verfahren verwendet oder modifiziert werden.

[0069] Aus den Koeffizienten können schliesslich durch eine Rücktransformation wieder das Original oder originalgetreu klingende Signale erzeugt werden, beispielsweise im Wave Format, welche dann beispielsweise über das Computermusiksystem und Lautsprecher wiedergegeben werden können. Im speziellen Fall können Klänge, welche durch Musiknoten oder Bilder irgendwelcher Art am Bildschirm dargestellt werden, aus den TF-Koeffizienten rekonstruiert und abgespielt werden.

[0070] Das erfindungsgemässe Verfahren bzw. die logische oder physische Verschaltung der Vorrichtung werden nachfolgend anhand der Ablauf- und Anordnungsbeziehungen der einzelnen Komponenten sowie der graphischen Darstellung auf einem Bildschirm beispielhaft und rein schematisch näher erläutert.

[0071] Im einzelnen zeigen

Fig.1 eine schematische Darstellung der einzelnen Schritte des erfindungsgemässen Verfahrens;

Fig.2 eine schematische Darstellung von Bereitstellungsalternativen für ein Eingangssignal;

Fig.3 eine schematische Darstellung der Zerlegung des Eingangssignals in Frequenzbereiche;

Fig.4 eine schematische Darstellung eines Transformierens der Frequenzbereiche;

Fig.5 eine schematische Darstellung der Schritte zur Notenerkennung durch harmonischen Dekomposition;

Fig.6 eine Darstellung einer graphischen Benutzeroberfläche zur interaktiven Bereitstellung von Zusatzinformationen;

- Fig.7 eine Darstellung eines ersten Schrittes in einem ersten Beispiel zur interaktiven Bereitstellung von Zusatzinformationen durch Setzen von Taktmarkierungen;
- Fig.8 eine Darstellung eines zweiten Schrittes in einem ersten Beispiel zur interaktiven Bereitstellung von Zusatzinformationen durch Setzen von Taktmarkierungen;
- Fig.9 eine Darstellung eines ersten Schrittes in einem zweiten Beispiel zur interaktiven Bereitstellung von Zusatzinformationen durch Anpassung des Verstärkungsfaktors und
- Fig.10 eine Darstellung eines zweiten Schrittes in einem zweiten Beispiel zur interaktiven Bereitstellung von Zusatzinformationen durch Anpassung des Verstärkungsfaktors.

[0072] Fig.1 zeigt eine schematische Darstellung der einzelnen Schritte des erfindungsgemässen Verfahrens.

[0073] Das akustische Signal wird durch eine Aufnahmekomponente erfasst oder von einem Datenträger importiert und in Form eines Eingangssignals ES zur Weiterverarbeitung bereitgestellt. Dieses Eingangssignal ES wird in einem Subband Coder SC in einzelne Frequenzbänder zerlegt, die nachfolgend jeweils einer frequenzoptimierten ersten Transformation TF1 und einer zeitoptimierten zweiten Transformation TF2 zugeführt werden. Diese Transformationsprozesse können parallel auch Informationen aus dem originalen Eingangssignal ES gewinnen und für den Transformationsprozess nutzen.

[0074] Die Ergebnisse der beiden Transformationen werden in einem Transformations-Prozessor TP - gegebenenfalls unter Rückkopplung mit der ersten Transformation TF1 und der zweiten Transformation TF2 - zu einer Koeffizientendatei zusammengeführt.

[0075] Auf der Basis dieser Koeffizientendatei erfolgt die harmonische Dekomposition HD zur Erkennung von dem Eingangssignal ES inhärenten Mustern. Dabei können zur harmonischen Dekomposition HD vorgegebene Koeffizienten genutzt werden, die beispielsweise in einem Speicher abgelegt sind oder über externe Datenträger zugeführt werden.

[0076] Die identifizierten Muster werden über eine graphische Konversion für eine graphische Schnittstelle exportierbar bzw. darstellbar gemacht. Ein Beispiel hierfür stellen die Umsetzung in Noten und beispielsweise der Ausdruck einer Partitur dar. Erfolgt eine Darstellung auf einer graphischen Benutzeroberfläche, so können interaktiv Parameter verändert oder vorgegeben werden sowie weitere Auswahlen oder Modifikationen erfolgen.

[0077] Zum Transfer von Dateien wird eine Schnittstelle EX/IM verwendet. Darüber hinaus kann nach einer Formatkonversion die akustische Darstellung der Muster über einen Audioausgang, der z.B. mit einem Synthesizer verbunden ist, erfolgen.

[0078] In Fig.2 erfolgt die schematische Darstellung von Bereitstellungsalternativen für das Eingangssignal ES. Das Eingangssignal kann durch verschiedenartige Quellen bereitgestellt werden. Hierzu gehören zeitnah oder in Echtzeit erfolgende Aufnahme wie auch die Verwendung gespeicherter Daten. Dabei können beispielsweise Signale im Wave-Format und Dateien von Audio-CDs direkt verwendet werden. Dateien in den Formaten MPx (MP3, MP4) oder WMA oder einem anderen Format werden zuerst durch Decoder in Wave-Dateien umgewandelt. Dazu stehen handelsübliche Funktionsbibliotheken, z.B. für MP3 vom Fraunhofer Institut, im Internet zur Verfügung. Alternativ können die Koeffizienten von MP3 oder vergleichbaren Formaten direkt oder über eine Vorbehandlung (z.B. Skalierung) in eine oder mehrere Schichten der Pyramidenzerlegung des Signals eingeordnet werden. Decoder für andere Formate, wie z.B. Ogg oder WMA, werden im Internet bereitgestellt, z.B. auf www.microsoft.com.

[0079] Ein Aufnahmepuffer AP ist Bestandteil eines Signalaufnahmeverfahrens auf dem Computer, beispielsweise DirectX der Firma Microsoft. Hierdurch können z.B. Aufnahmen von Signalen über ein an den Computer angeschlossenes Mikrofon erfolgen.

[0080] Die Zerlegung des Eingangssignals ES in Frequenzbereiche im Subband Coder SC wird in Fig.3 schematisch dargestellt.

[0081] Das als Wave-Datei bereitgestellte Eingangssignal ES wird durch geeignete Hochpassfilter HP und Tiefpassfilter TP und durch Verminderung der Samplingrate, z.B. durch eine Halbierung der Datenrate HDR, in Unterbereiche oder Subbänder SBB zerteilt. Typischerweise enthält jedes Subband SBB eine bandpass-gefilterte Version des Eingangssignals ES. Beispiele für Filterkerne sind

- o für Tiefpass {0.25, 0.5, 0.25} oder {0.05, 0.2, 0.4, 0.2, 0.05} und
- o für Hochpässe Filterkerne deren Mittelwert der Koeffizienten Null (0.0) ergibt, z.B. {-1, 2, -1}.

[0082] Wahlweise können die Hochpassfilter auch weggelassen werden, wodurch sich eine Serie von tiefpassgefilterten Subbändern erzeugen lässt.

[0083] Fig.4 veranschaulicht das Transformieren der Frequenzbereiche in einer schematischen Darstellung. Die einzelnen Subbänder SBB werden den beiden unterschiedlich optimierten Transformationen TF1 und TF2 unterworfen

und nachfolgend in verschiedenen Schichten TFL0, TFL1,...TFLN gespeichert. Das in den Schichten TFL0, TFL1,...TFLN der Ausgangsmenge gespeicherte Signal ist beispielsweise eine Mischung aus den Transformationsausgangssignalen und einer Pyramidenzerlegung der jeweils nächsthöheren Stufe der Pyramide. Je nach konkreten Anwendungszweck und zu verarbeitenden Typen von akustischen Eingangssignalen ES kann auch eine andere Zerlegungsart oder auch

eine mehrfache Pyramidenzerlegung erfolgen.

[0084] Fig.5 zeigt eine schematische Darstellung der Schritte zur Notenerkennung durch harmonischen Dekomposition HD. Die in den verschiedenen Schichten TFL0, TFL1,...TFLN enthaltenen Informationen werden in einem Filter FI zusammengefasst und danach zur Ereignisextraktion der harmonischen Dekomposition unterworfen, in denen die Mustererkennung und Modellbildung stattfindet. Hierzu können erfindungsgemäss eine Vielzahl von vorstehend beschriebenen Ansätzen verwendet werden. Die Ergebnisse der harmonischen Dekomposition HD werden beispielsweise graphisch in Form von Noten dargestellt, so dass durch einen Benutzer oder auch andere Verfahren eine Auswahl oder Vorgabe von Informationen erfolgen kann, die wiederum Eingang in den Schritt der harmonischen Dekomposition HD finden.

[0085] Ein Beispiel für eine graphische Benutzeroberfläche zur interaktiven Bereitstellung von Zusatzinformationen wird in Fig.6 dargestellt. Die Oberfläche stellt unter anderem einen Verstärkungsregler 1 und einen manuell veränderbaren Taktmarkierer 2 zur Festsetzung von Takten bereit.

[0086] Die Anwendung des Taktmarkierers 2 wird in Fig. 7 in einem ersten Schritt eines ersten Beispiels zur interaktiven Bereitstellung von Zusatzinformationen durch Setzen von Taktmarkierungen erläutert. Dieser Ansatz erlaubt eine Bestimmung aller Takte im gesamten Lied. Durch den Taktmarkierer 2 wird ein Takt im Lied identifiziert und graphisch durch eine Raute 3 in der obersten Zeile angezeigt. Das Betätigen eines Funktionselementes führt dann zum Umrechnen der Ereignisse in standardgemässe Musiknoten, wobei die automatisch gesetzten Takte durch Dreiecke 4 in der obersten Zeile markiert werden. Verbesserungen dieser Methode können noch dadurch erreicht werden, dass die Tonsuren, speziell dabei die Schlagzeugspur, zur Feinabstimmung der Takte herangezogen werden können. Trotzdem kann es aufgrund von Variationen in der gespielten Musik, durch Schwankungen der Aufnahmegeschwindigkeit oder Driteffekte zu einem Auseinanderfallen von errechneten Takten und tatsächlichen Mustern in der Aufnahme kommen, wie im Beispiel innerhalb des gestrichelten Bereichs durch Pfeile angezeigt.

[0087] Durch das manuelle Anpassen der Taktmarkierung kann dieses Auseinanderfallen wieder korrigiert werden, wie in Fig.8 gezeigt.

[0088] In Fig.9 erfolgt eine Darstellung eines ersten Schrittes in einem zweiten Beispiel zur interaktiven Bereitstellung von Zusatzinformationen durch Anpassung des Verstärkungsfaktors. In diesem Beispiel wird der Schwellwertregler mit einem Schwellwert grösser 0 gewählt, so dass nur Notenereignisse angezeigt werden, welche grösser als der Schwellwert sind. Einige relevante Bereiche werden durch Ellipsen markiert.

[0089] In diesen Bereichen werden nach Veränderung der Einstellung des Schwellwertreglers weitere Informationen sichtbar, wie in Fig.10 gezeigt. Wird der Schwellwertregler auf Null gesetzt, so werden alle Notenereignisse sichtbar und es werden alle ermittelten Ereignisse angezeigt. Durch die Variation des Schwellwertreglers können somit Anpassungen des Ergebnisses erfolgen, ohne dass das gesamte Verfahren von Anfang an neu durchgeführt werden muss.

Patentansprüche

1. Verfahren zur Musterzuordnung für akustischen Aufnahmen mit den Schritten

- Bereitstellen eines Signals welches eine akustische Aufnahme repräsentiert;
- Zerlegen des Signals in Frequenzbereiche;
- Transformieren der Frequenzbereiche zur spektralen Zerlegung in wenigstens eine Koeffizientendatei;
- Durchführen einer harmonischen Dekomposition der Koeffizientendatei; und
- Musterzuordnung;

dadurch gekennzeichnet, dass

beim Transformieren der Frequenzbereiche, insbesondere jeweils für alle Frequenzbereiche, wenigstens

- eine hinsichtlich der Frequenzauflösung optimierte erste Transformation und
- eine hinsichtlich der Zeitauflösung optimierte zweite Transformation erfolgt.

2. Verfahren nach Anspruch 1,

dadurch gekennzeichnet, dass

beim Transformieren der Frequenzbereiche eine optimierte Selektion der Koeffizienten aus den Ergebnissen der ersten Transformation und der zweiten Transformation und/oder eine Mischung der Koeffizienten aus den Ergeb-

nissen der ersten Transformation und der zweiten Transformation erfolgt.

3. Verfahren nach Anspruch 2,
dadurch gekennzeichnet, dass
beim Transformieren der Frequenzbereiche

- die erste Transformation mit einem längeren Zeitfenster und
- die zweite Transformation mit einem kürzeren Zeitfenster erfolgt,

insbesondere wobei die Selektion anhand des Verhältnisses der Realteile von erster und zweiter Transformation getroffen wird.

4. Verfahren nach Anspruch 2,
dadurch gekennzeichnet, dass
beim Transformieren der Frequenzbereiche die Selektion oder Mischung anhand der frequenzabhängigen Steigung des Phasensignals jeweils für die Ergebnisse der ersten Transformation und der zweiten Transformation erfolgt.

5. Verfahren nach Anspruch 2,
dadurch gekennzeichnet, dass
beim Transformieren der Frequenzbereiche die Selektion oder Mischung anhand eines Vergleichs der Ergebnisse der ersten Transformation und der zweiten Transformation mit einem Satz von vorgegebenen Koeffizienten erfolgt.

6. Verfahren nach einem der vorangehenden Ansprüche,
dadurch gekennzeichnet, dass
die erste Transformation und/oder zweite Transformation nach einem der folgenden Prinzipien erfolgt

- Diskrete Fourier-Transformation,
- Fast-Fourier-Transformation,
- Wavelet-Transformation,
- Sinus Transformation,
- Kosinus Transformation.

7. Verfahren nach einem der vorangehenden Ansprüche,
dadurch gekennzeichnet, dass
beim Transformieren der Frequenzbereiche für jede Transformation ein Aggregat der Ergebnisse, insbesondere das zeitliche Integral für eine Frequenz, berücksichtigt werden.

8. Verfahren nach einem der vorangehenden Ansprüche,
dadurch gekennzeichnet, dass
das Zerlegen des Signals nach dem Prinzip der Teilung in Oktaven und/oder der Pyramidenzerlegung erfolgt.

9. Verfahren nach einem der vorangehenden Ansprüche,
dadurch gekennzeichnet, dass
beim Durchführen der harmonischen Dekomposition
- ein Vergleich mit vorgegebenen Koeffizienten, insbesondere unter Minimierung des Residuums, oder
 - ein Vergleich mit Koeffizienten aus einer vorangegangenen Analyse des Signals, insbesondere durch Ableiten von Koeffizienten unter Verwendung eines charakteristischen Grundprofils,

erfolgt.

10. Verfahren nach einem der vorangehenden Ansprüche,
dadurch gekennzeichnet, dass
beim Durchführen der harmonischen Dekomposition eine Interaktion mit einem Benutzer erfolgt, insbesondere durch Eingabe von Zusatzinformationen.

11. Verfahren nach einem der vorangehenden Ansprüche,
dadurch gekennzeichnet, dass

beim Durchführen der harmonischen Dekomposition originale und/oder synthetische Frequenzanteile, insbesondere obere Frequenzanteile, verwendet werden.

5 **12.** Computerprogrammprodukt mit Programmcode, der auf einem maschinenlesbaren Träger gespeichert oder durch eine elektromagnetische Welle verkörpert ist, zur Durchführung des Verfahrens nach einem der Ansprüche 1 bis 11.

13. Vorrichtung zur Durchführung eines Verfahrens nach einem der Ansprüche 1 bis 11 mit wenigstens

- 10
- einer Aufnahmekomponente zur Aufnahmen eines akustischen Signals,
 - einem Subband-Coder zum Zerlegen des Signals in einzelne Frequenzbereiche,
 - einem Transformations-Prozessor zur spektralen Zerlegung der Frequenzbereiche in wenigstens eine Koeffizientendatei,
 - 15 - einer Exportschnittstelle zum Export der Koeffizientendatei,

dadurch gekennzeichnet, dass

dem Transformations-Prozessor eine erste Transformationsstufe und eine zweite Transformationsstufe zugeordnet sind, wobei die erste Transformationsstufe eine optimierte Frequenzauflösung und die zweite Transformationsstufe eine optimierte Zeitauflösung bewirkt.

20 **14.** Koeffizientendatei zur Verwendung in einem Verfahrens nach einem der Ansprüche 1 bis 11

gekennzeichnet durch

die Koeffizienten der spektralen Zerlegung des akustischen Signals und zugeordneter Informationen zur Signalstatistik.

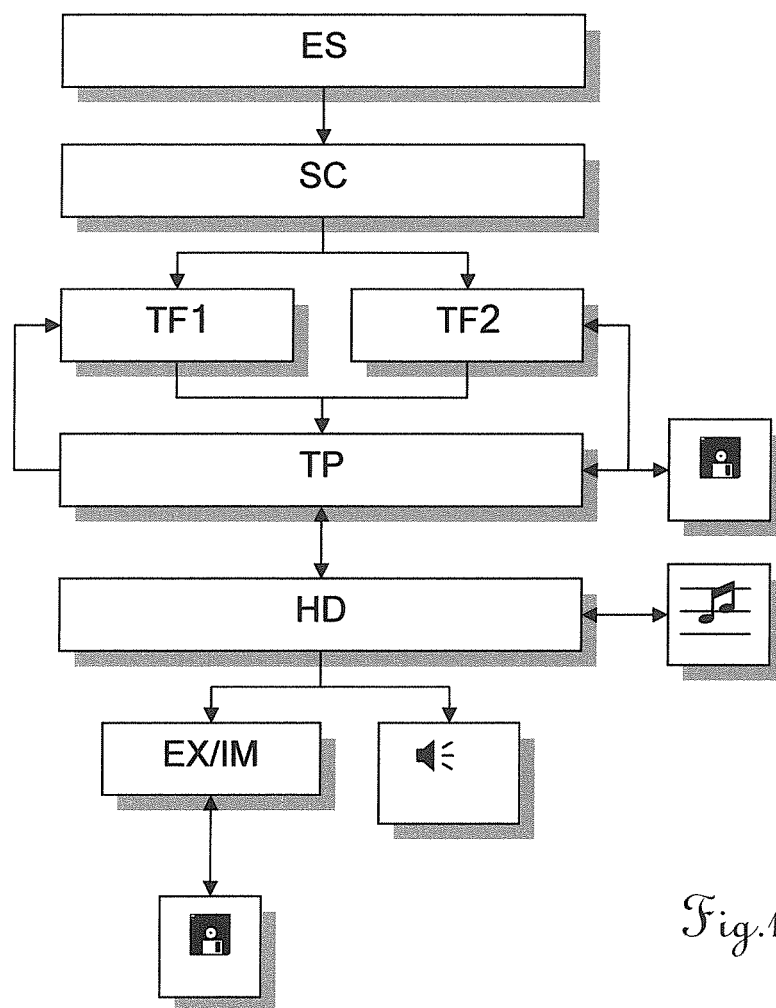


Fig.1

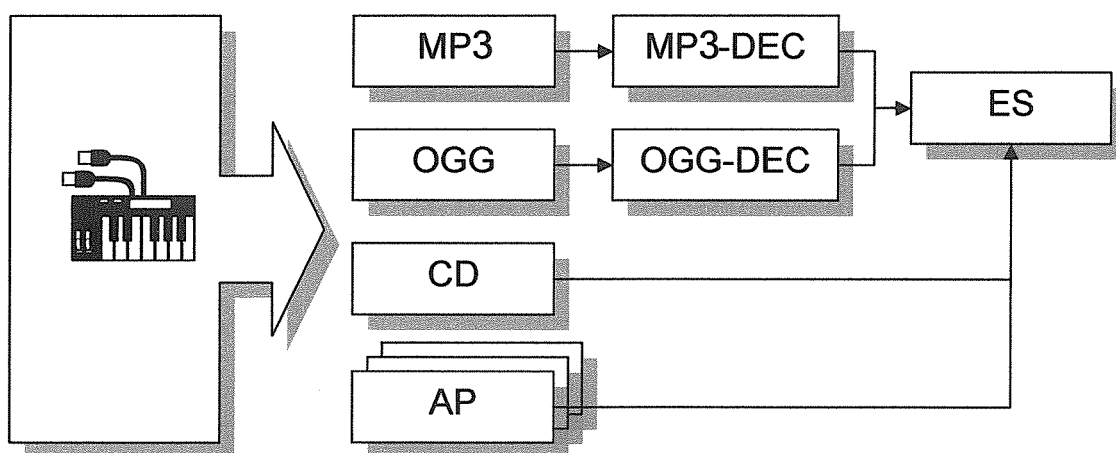


Fig.2

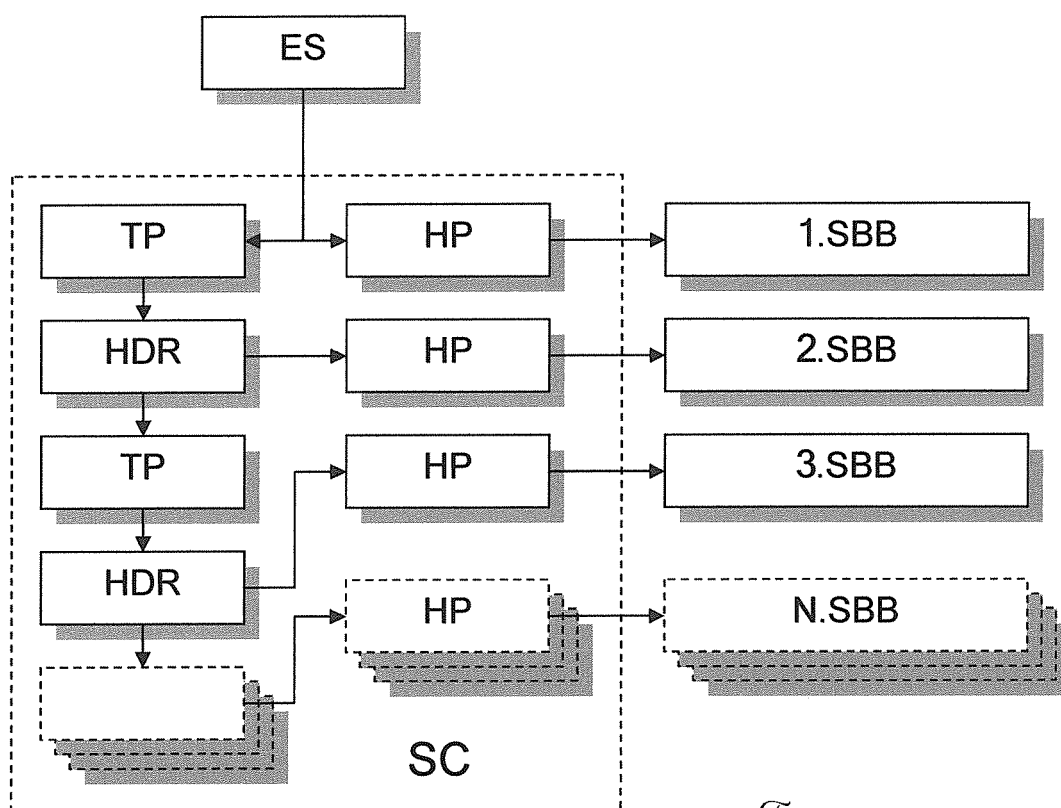


Fig.3

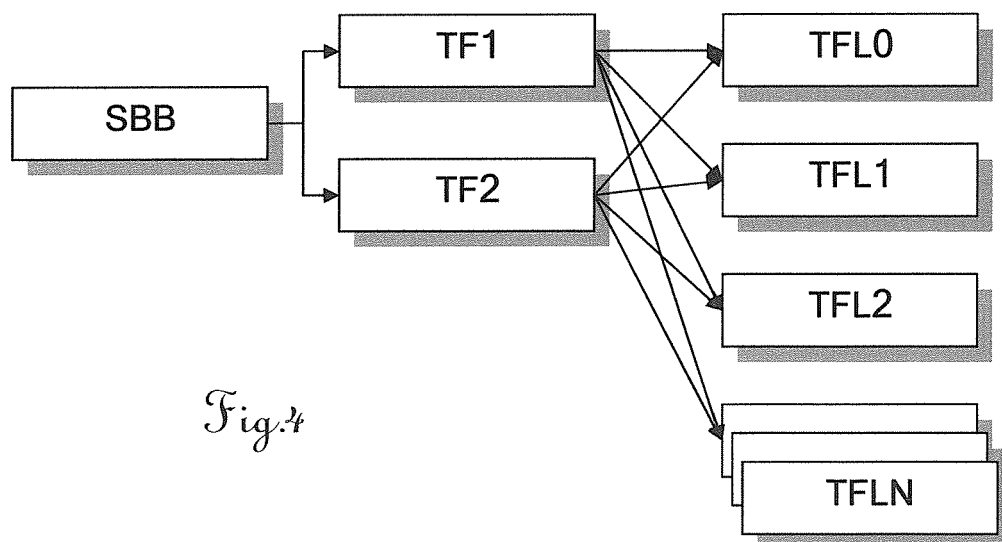


Fig.4

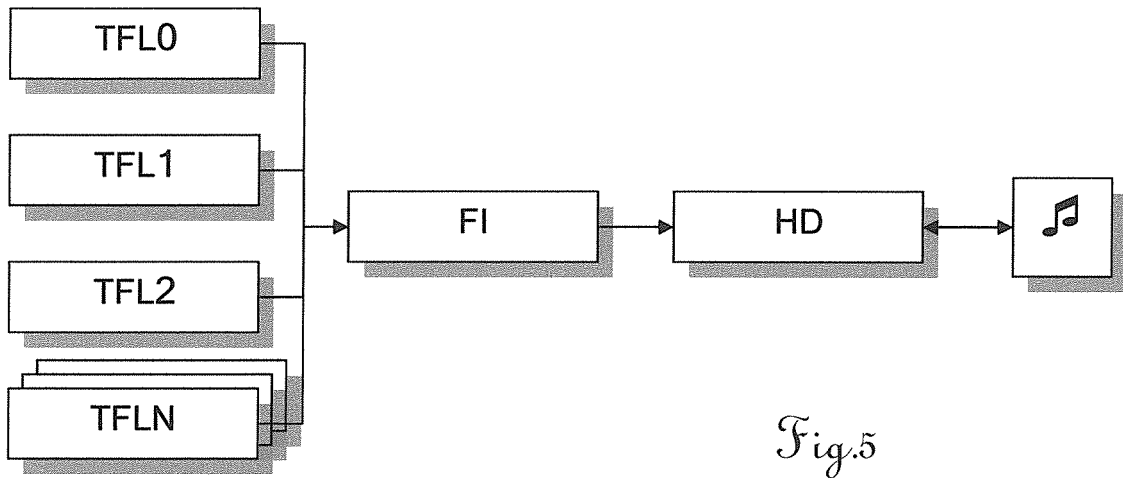


Fig.5

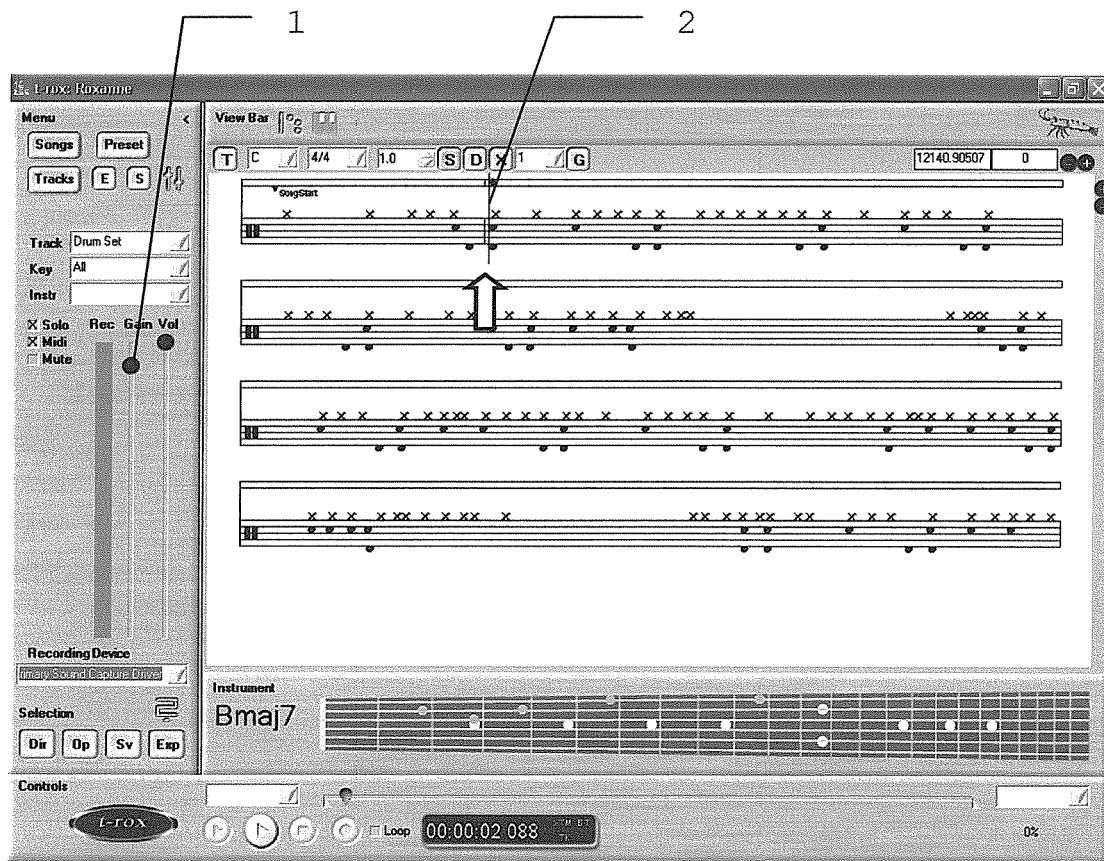


Fig.6

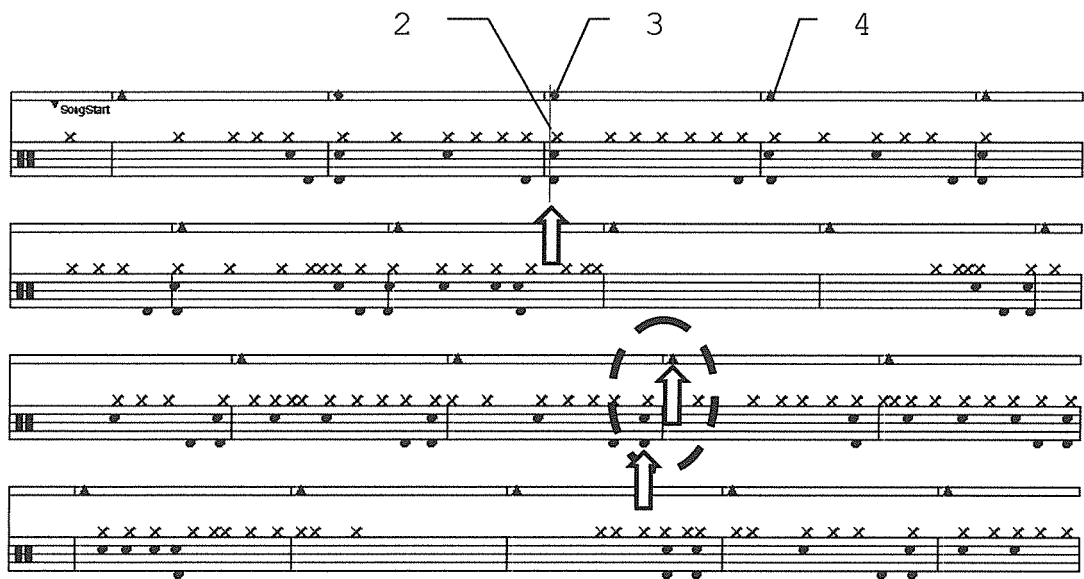


Fig.7

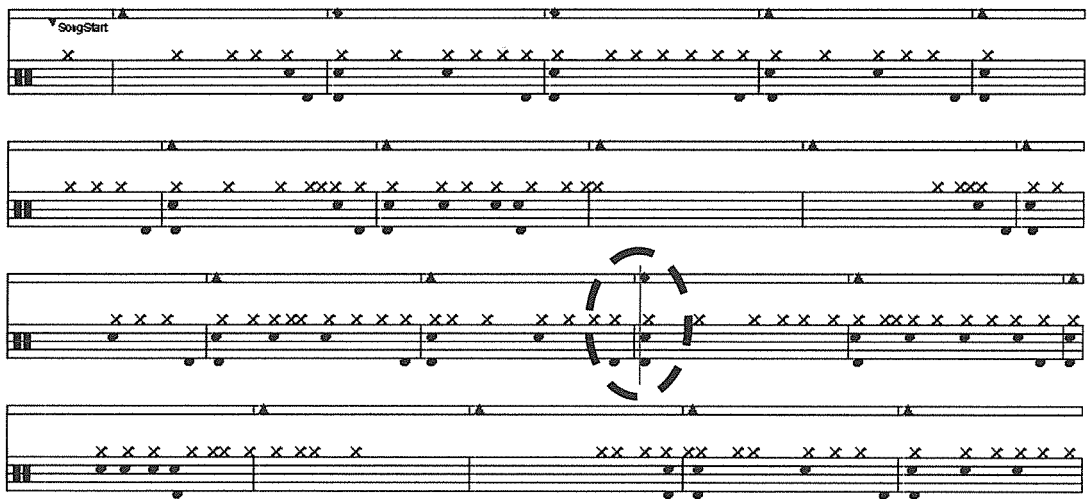


Fig.8

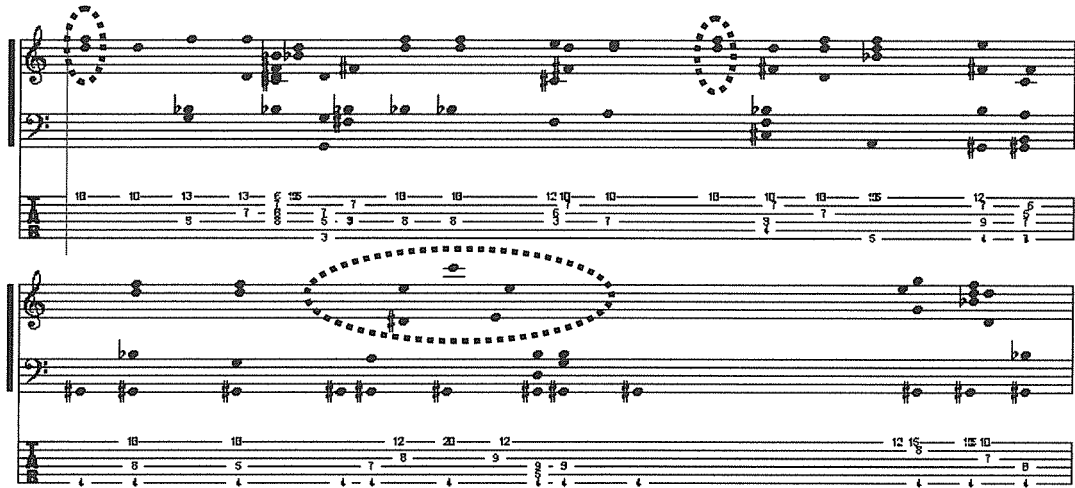


Fig.9



Fig.10



Europäisches
Patentamt

EUROPÄISCHER RECHERCHENBERICHT

Nummer der Anmeldung
EP 05 10 7730

EINSCHLÄGIGE DOKUMENTE			
Kategorie	Kennzeichnung des Dokuments mit Angabe, soweit erforderlich, der maßgeblichen Teile	Betrifft Anspruch	KLASSIFIKATION DER ANMELDUNG (IPC)
Y	MELIH K ET AL: "Audio retrieval using perceptually based structures" MULTIMEDIA COMPUTING AND SYSTEMS, 1998. PROCEEDINGS. IEEE INTERNATIONAL CONFERENCE ON AUSTIN, TX, USA 28 JUNE-1 JULY 1998, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 28. Juni 1998 (1998-06-28), Seiten 338-347, XP010291569 ISBN: 0-8186-8557-3 1. Introduction 4.1 Overview 4.2 TFD generation 4.1 (sic) Masking thresholds 4.2 (sic) Peak icking and tracking 4.4 Track classification and segregation * Abbildungen 2-10 *	1-14	G10L11/00 G10H1/00
Y	BOSI M ET AL: "ISO/IEC MPEG-2 ADVANCED AUDIO CODING" JOURNAL OF THE AUDIO ENGINEERING SOCIETY, AUDIO ENGINEERING SOCIETY, NEW YORK, NY, US, Bd. 45, Nr. 10, Oktober 1997 (1997-10), Seiten 789-812, XP000730161 ISSN: 1549-4950 3. Filter Bank * Abbildungen 3,4,6,7 *	1-14	RECHERCHIERTE SACHGEBIETE (IPC) G10L G10H G10G
Der vorliegende Recherchenbericht wurde für alle Patentansprüche erstellt			
Recherchenort München		Abschlußdatum der Recherche 19. Januar 2006	Prüfer Lecoainte, M
KATEGORIE DER GENANNTEN DOKUMENTE X : von besonderer Bedeutung allein betrachtet Y : von besonderer Bedeutung in Verbindung mit einer anderen Veröffentlichung derselben Kategorie A : technologischer Hintergrund O : nichtschriftliche Offenbarung P : Zwischenliteratur		T : der Erfindung zugrunde liegende Theorien oder Grundsätze E : älteres Patentdokument, das jedoch erst am oder nach dem Anmeldedatum veröffentlicht worden ist D : in der Anmeldung angeführtes Dokument L : aus anderen Gründen angeführtes Dokument & : Mitglied der gleichen Patentfamilie, übereinstimmendes Dokument	

1
EPO FORM 1503 03.82 (P04C03)



Europäisches
Patentamt

EUROPÄISCHER RECHERCHENBERICHT

Nummer der Anmeldung
EP 05 10 7730

EINSCHLÄGIGE DOKUMENTE			
Kategorie	Kennzeichnung des Dokuments mit Angabe, soweit erforderlich, der maßgeblichen Teile	Betrifft Anspruch	KLASSIFIKATION DER ANMELDUNG (IPC)
Y	EDLER B: "CODIERUNG VON AUDIOSIGNALEN MIT UEBERLAPPENDER TRANSFORMATION UND ADAPTIVEN FENSTERFUNKTIONEN CODING OF AUDIO SIGNALS WITH OVERLAPPING BLOCK TRANSFORM AND ADAPTIVE WINDOW FUNCTIONS" FREQUENZ, SCHIELE UND SCHON, BERLIN, DE, Bd. 43, Nr. 9, 1. September 1989 (1989-09-01), Seiten 252-256, XP000052987 ISSN: 0016-1136 3. OBT mit adaptiven Fensterfunktionen 4. Steuerung der Fensteradaptation -----	1-14	
A	YASUI NAGOSHI, YUSUKE HIOKA, NOZOMU HAMADA: "Transcription of Music Composed of Melody and Chord Using Tree-structured Filter Bank" PROCEEDINGS OF THE IASTED INTERNATIONAL CONFERENCE, 16. August 2001 (2001-08-16), Seiten 415-419, XP009059920 Honolulu, Hawaii, USA 1. Introduction 2. Method Proposed * Zusammenfassung; Abbildungen 2-5 * -----	1-14	RECHERCHIERTE SACHGEBIETE (IPC)
D,A	US 6 140 568 A (KOHLE ET AL) 31. Oktober 2000 (2000-10-31) * Zusammenfassung * * Spalte 3, Zeile 1 - Spalte 7, Zeile 14 * * Abbildungen 1,2 * -----	1,12-14	
A	US 5 752 224 A (TSUTSUI ET AL) 12. Mai 1998 (1998-05-12) * Spalte 3 - Spalte 6 * * Abbildungen 1-5 * -----	1,12-14	
Der vorliegende Recherchenbericht wurde für alle Patentansprüche erstellt			
Recherchenort München		Abschlußdatum der Recherche 19. Januar 2006	Prüfer Lecoite, M
KATEGORIE DER GENANNTEN DOKUMENTE X : von besonderer Bedeutung allein betrachtet Y : von besonderer Bedeutung in Verbindung mit einer anderen Veröffentlichung derselben Kategorie A : technologischer Hintergrund O : nichtschriftliche Offenbarung P : Zwischenliteratur		T : der Erfindung zugrunde liegende Theorien oder Grundsätze E : älteres Patentdokument, das jedoch erst am oder nach dem Anmeldedatum veröffentlicht worden ist D : in der Anmeldung angeführtes Dokument L : aus anderen Gründen angeführtes Dokument & : Mitglied der gleichen Patentfamilie, übereinstimmendes Dokument	

1
EPO FORM 1503 03.82 (P04C03)

**ANHANG ZUM EUROPÄISCHEN RECHERCHENBERICHT
 ÜBER DIE EUROPÄISCHE PATENTANMELDUNG NR.**

EP 05 10 7730

In diesem Anhang sind die Mitglieder der Patentfamilien der im obengenannten europäischen Recherchenbericht angeführten Patentdokumente angegeben.
 Die Angaben über die Familienmitglieder entsprechen dem Stand der Datei des Europäischen Patentamts am
 Diese Angaben dienen nur zur Unterrichtung und erfolgen ohne Gewähr.

19-01-2006

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 6140568 A	31-10-2000	KEINE	
US 5752224 A	12-05-1998	KEINE	

EPO FORM P0461

Für nähere Einzelheiten zu diesem Anhang : siehe Amtsblatt des Europäischen Patentamts, Nr.12/82

IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

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

In der Beschreibung aufgeführte Patentdokumente

- US 6140568 A [0007]
- US 6355869 B1 [0008]