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

(11) EP 1 377 960 B1

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
24.11.2004 Patentblatt 2004/48

(21) Anmeldenummer: **02730100.1**

(22) Anmeldetag: **04.04.2002**

(51) Int Cl.[7]: **G10H 1/00**

(86) Internationale Anmeldenummer:
PCT/EP2002/003736

(87) Internationale Veröffentlichungsnummer:
WO 2002/084641 (24.10.2002 Gazette 2002/43)

(54) **VERFAHREN ZUM ÜBERFÜHREN EINES MUSIKSIGNALS IN EINE NOTEN-BASIERTE BESCHREIBUNG UND ZUM REFERENZIEREN EINES MUSIKSIGNALS IN EINER DATENBANK**

METHOD FOR CONVERTING A MUSIC SIGNAL INTO A NOTE-BASED DESCRIPTION AND FOR REFERENCING A MUSIC SIGNAL IN A DATA BANK

PROCEDE POUR CONVERTIR UN SIGNAL MUSICAL EN UNE DESCRIPTION FONDEE SUR DES NOTES ET POUR REFERENCER UN SIGNAL MUSICAL DANS UNE BASE DE DONNEES

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

(30) Priorität: **10.04.2001 DE 10117870**

(43) Veröffentlichungstag der Anmeldung:
07.01.2004 Patentblatt 2004/02

(73) Patentinhaber: **Fraunhofer-Gesellschaft für angewandte Forschung e.V.**
80636 München (DE)

(72) Erfinder:
- **KLEFENZ, Frank**
  68159 Mannheim (DE)
- **BRANDENBURG, Karlheinz**
  91054 Erlangen (DE)
- **KAUFMANN, Matthias**
  98693 Ilmenau (DE)

(74) Vertreter: **Zimmermann, Tankred Klaus**
Schoppe, Zimmermann, Stöckeler & Zinkler
Patentanwälte
Postfach 246
82043 Pullach bei München (DE)

(56) Entgegenhaltungen:
EP-A- 0 944 033
WO-A-01/69575
US-A- 5 210 820
US-A- 5 874 686

EP 1 377 960 B1

Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf das Gebiet der Verarbeitung von Musiksignalen und insbesondere auf das Umsetzen eines Musiksignals in eine Noten-basierte Beschreibung.

[0002] Konzepte, mit denen Lieder durch Vorgabe einer Tonfolge referenziert werden, sind für viele Anwender nützlich. Wer kennt nicht die Situation, daß man die Melodie eines Liedes vor sich her singt, sich aber außer der Melodie nicht an den Titel des Liedes erinnern kann. Wünschenswert wäre, eine Melodiesequenz vorzusingen oder mit einem Musikinstrument vorzuspielen, und mit diesen Informationen die Melodiesequenz in einer Musikdatenbank zu referenzieren, wenn die Melodiesequenz in der Musikdatenbank enthalten ist.

[0003] Eine standardmäßige Noten-basierte Beschreibung von Musiksignalen ist das MIDI-Format (MIDI = Music Interface Description). Eine MIDI-Datei umfaßt eine Noten-basierte Beschreibung derart, daß der Tonanfang und das Tonende eines Tons bzw. der Tonanfang und die Dauer des Tons als Funktion der Zeit aufgezeichnet sind. MIDI-Dateien können beispielsweise in elektronischen Keyboards eingelesen werden und "abgespielt" werden. Selbstverständlich existieren auch Soundkarten zum Abspielen eines MIDI-Files über die mit der Soundkarte eines Computers verbundenen Lautsprecher. Daraus ist zu sehen, daß das Umformen einer Noten-basierten Beschreibung, welches in seiner ursprünglichsten Form durch einen Intrumentalisten "manuell" durchgeführt wird, der ein durch Noten aufgezeichnetes Lied mittels eines Musikinstruments spielt, auch ohne weiteres automatisch durchgeführt werden kann.

[0004] Das Gegenteil ist jedoch ungleich aufwendiger. Die Umformung eines Musiksignals, das eine gesungene Melodiesequenz, eine gespielte Melodiesequenz, eine von einem Lautsprecher aufgezeichnete Melodiesequenz oder eine in Form einer Datei vorhandene digitalisierte und optional komprimierte Melodiesequenz ist, in eine Noten-basierte Beschreibung in Form einer MIDI-Datei oder in eine konventionelle Notenschrift ist mit großen Einschränkungen verbunden.

[0005] In der Dissertation "Using Contour as a Mid-Level Representation of Melody" von A. Lindsay, Massachusetts Institute of Technology, September 1996, ist ein Verfahren zum Umformen eines gesungenen Musiksignals in eine Folge von Noten beschrieben. Ein Lied muß unter Verwendung von Stoppkonsonanten vorgetragen werden, d. h. als eine Folge von "da", "da", "da". Anschließend wird die Leistungsverteilung des von dem Sänger erzeugten Musiksignals über der Zeit betrachtet. Aufgrund der Stoppkonsonanten ist zwischen dem Ende eines Tons und dem Beginn des darauffolgenden Tons ein deutlicher Leistungseinbruch in einem Leistungs-Zeit-Diagramm zu erkennen. Auf der Basis der Leistungseinbrüche wird eine Segmentierung des Musiksignals durchgeführt, so daß in jedem Segment eine Note vorhanden ist. Eine Frequenzanalyse liefert die Höhe des gesungenen Tons in jedem Segment, wobei die Folge von Frequenzen auch als Pitch-Contourlinie bezeichnet wird.

[0006] Das Verfahren ist dahingehend nachteilig, daß es auf eine gesungene Eingabe beschränkt ist. Als Vorgabe muß die Melodie durch einen Stoppkonsonanten und einen Vokalpart gesungen werden, in der Form "da" "da" "da", damit eine Segmentierung des aufgezeichneten Musiksignals vorgenommen werden kann. Dies schließt bereits eine Anwendung des Verfahrens auf Orchesterstücke aus, in denen ein dominantes Instrument gebundenen Noten, d. h. nicht durch Pausen getrennte Noten, spielt.

[0007] Nach einer Segmentierung berechnet das bekannte Verfahren Intervalle jeweils zwei aufeinanderfolgender Pitch-Werte, d. h. Tonhöhenwerte, in der Pitchwertfolge. Dieser Intervallwert wird als Abstandsmaß angenommen. Die sich ergebende Pitchfolge wird dann mit in einer Datenbank gespeicherten Referenzfolgen verglichen, wobei das Minimum einer Summe quadrierter Differenzbeträge über alle Referenzfolgen als Lösung, d. h. als in der Datenbank referenzierte Notenfolge, angenommen wird.

[0008] Ein weiterer Nachteil dieses Verfahrens besteht darin, daß ein Pitch-Tracker eingesetzt wird, welcher Oktav-Sprungfehler aufweist, die nachträglich kompensiert werden müssen. Ferner muß der Pitch-Tracker fein abgestimmt werden, um gültige Werte zu liefern. Das Verfahren nutzt lediglich die Intervallabstände zweier aufeinanderfolgender Pitch-Werte. Eine Grobquantisierung der Intervalle wird durchgeführt, wobei diese Grobquantisierung lediglich grobe Schritte aufweist, die als "sehr groß", "groß", "gleichbleibend" eingeteilt sind. Durch diese Grobquantisierung gehen die absoluten Tonangaben in Hertz verloren, wodurch eine feinere Bestimmung der Melodie nicht mehr möglich ist.

[0009] Um eine Musikerkennung durchführen zu können, ist es wünschenswert, aus einer gespielten Tonfolge eine Noten-basierte Beschreibung beispielsweise in Form eines MIDI-Files oder in Form einer konventionellen Notenschrift zu bestimmen, wobei jede Note durch Tonanfang, Tonlänge und Tonhöhe gegeben ist.

[0010] Ferner ist zu bedenken, daß die Eingabe nicht immer exakt ist. Insbesondere für eine kommerzielle Nutzung muß davon ausgegangen werden, daß die gesungene Notenfolge sowohl hinsichtlich der Tonhöhe als auch hinsichtlich des Tonrhythmus und der Tonfolge unvollständig sein kann. Wenn die Notenfolge mit einem Instrument vorgespielt werden soll, so muß davon ausgegangen werden, daß das Instrument unter Umständen verstimmt ist, auf einen anderen Frequenzgrundton gestimmt ist (beispielsweise nicht auf den Kammerton A von 440 Hz sondern auf das "A" bei 435 Hz). Ferner kann das Instrument in einer eigenen Tonart gestimmt sein, wie z. B. die B-Klarinette oder das Es-Saxophon. Die Melodietonfolge kann auch bei instrumentaler Dar-

bietung unvollständig sein, indem Töne weggelassen sind (Delete), indem Töne eingestreut sind (Insert), oder indem andere (falsche) Töne gespielt werden (Replace). Ebenso kann das Tempo variiert sein. Weiterhin ist zu berücksichtigen, daß jedes Instrument eine eigene Klangfarbe aufweist, so daß ein von einem Instrument gespielter Ton eine Mischung aus Grundton und anderen Frequenzanteilen, den sogenannten Obertönen, ist.

[0011] Die Aufgabe der vorliegenden Erfindung besteht darin, ein robusteres Verfahren und eine robustere Vorrichtung zum Überführen eines Musiksignals in eine Noten-basierte Beschreibung zu schaffen.

[0012] Diese Aufgabe wird durch ein Verfahren gemäß Patentanspruch 1 oder durch eine Vorrichtung gemäß Patentanspruch 31 gelöst.

[0013] Eine weitere Aufgabe der vorliegenden Erfindung besteht darin, ein robusteres Verfahren und eine robustere Vorrichtung zum Referenzieren eines Musiksignals in einer Datenbank, die eine Noten-basierte Beschreibung einer Mehrzahl von Datenbank-Musiksignalen aufweist, zu schaffen.

[0014] Diese Aufgabe wird durch ein Verfahren nach Patentanspruch 23 oder durch eine Vorrichtung nach Patentanspruch 32 gelöst.

[0015] Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, daß für eine effiziente und robuste Überführung eines Musiksignals in eine Noten-basierte Beschreibung eine Einschränkung dahingehend nicht akzeptabel ist, daß eine gesungene oder gespielte Notenfolge durch Stoppkonsonanten dargeboten werden muß, die dazu führen, daß die Leistungs-Zeit-Darstellung des Musiksignals scharfe Leistungseinbrüche aufweist, welche dazu verwendet werden können, eine Segmentierung des Musiksignals durchzuführen, um einzelne Töne der Melodiefolge voneinander abgrenzen zu können.

[0016] Erfindungsgemäß wird aus dem vorgesungenen oder vorgespielten oder in einer sonstigen Form vorliegenden Musiksignal eine Noten-basierte Beschreibung dadurch gewonnen, daß zunächst eine Frequenz-Zeit-Darstellung des Musiksignals erzeugt wird, wobei die Frequenz-Zeit-Darstellung Koordinatentupel aufweist, wobei ein Koordinatentupel einen Frequenzwert und einen Zeitwert aufweist, wobei der Zeitwert die Zeit des Auftretens des zugeordneten Frequenz in dem Musiksignal angibt. Anschließend wird eine Fitfunktion als Funktion der Zeit berechnet, deren Verlauf durch die Koordinatentupel der Frequenz-Zeit-Darstellung bestimmt ist. Aus der Fitfunktion werden zumindest zwei benachbarte Extremwerte ermittelt. Die zeitliche Segmentierung der Frequenz-Zeit-Darstellung, um Töne einer Melodiefolge voneinander abgrenzen zu können, wird auf der Basis der ermittelten Extremwerte durchgeführt, wobei ein Segment durch die zumindest zwei benachbarten Extremwerte der Fitfunktion begrenzt ist, wobei die zeitliche Länge des Segments auf eine zeitliche Länge einer Note für das Segment hinweist. Damit wird ein Notenrhythmus erhalten. Die Notenhöhen werden schließlich unter Verwendung lediglich von Koordinaten-Tupeln in jedem Segment bestimmt, so daß für jedes Segment ein Ton ermittelt wird, wobei die Töne in den aufeinanderfolgenden Segmenten auf die Melodiefolge hinweisen.

[0017] Ein Vorteil der vorliegenden Erfindung besteht darin, daß eine Segmentierung des Musiksignals unabhängig davon erreicht wird, ob das Musiksignal von einem Instrument gespielt wird oder vorgesungen wird. Erfindungsgemäß ist es nicht mehr erforderlich, daß ein zu verarbeitendes Musiksignal einen Leistungs-Zeit-Verlauf hat, der scharfe Einbrüche aufweisen muß, um die Segmentierung vornehmen zu können. Die Eingabeart ist bei dem erfindungsgemäßen Verfahren somit nicht mehr beschränkt. Während das erfindungsgemäße Verfahren bei monophonen Musiksignalen, wie sie durch eine einzelne Stimme oder durch ein einzelnes Instrument erzeugt werden, am besten funktioniert, ist es auch für eine polyphone Darbietung geeignet, wenn in der polyphonen Darbietung ein Instrument bzw. eine Stimme vorherrschend ist.

[0018] Aufgrund der Tatsache, daß die zeitliche Segmentierung der Noten der Melodiefolge, die das Musiksignal darstellt, nicht mehr durch Leistungsbetrachtungen durchgeführt wird, sondern durch Berechnen einer Fitfunktion unter Verwendung einer Frequenz-Zeit-Darstellung, ist eine kontinuierliche Eingabe möglich, wie sie einem natürlichen Gesang oder einem natürlichen Instrumentenspiel am ehesten entspricht.

[0019] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird eine Instrumenten-spezifische Nachbearbeitung der Frequenz-Zeit-Darstellung durchgeführt, um die Frequenz-Zeit-Darstellung unter Kenntnis der Charakteristika eines bestimmten Instruments nachzubearbeiten, um eine genauere Pitch-Contour-Linie und damit eine genauere Tonhöhenbestimmung zu erreichen.

[0020] Ein Vorteil der vorliegenden Erfindung besteht darin, daß das Musiksignal von jedem beliebigen Harmonic-Sustained Musikinstrument vorgetragen werden kann, wobei zu den Harmonic-Sustained-Musikinstrumenten die Blechinstrumente, die Holzblasinstrumente oder auch die Saiteninstrumente, wie z. B. Zupfinstrumente, Streichinstrumente oder Anschlaginstrumente, zählen. Aus der Frequenz-Zeit-Verteilung wird unabhängig von der Klangfarbe des Instrumentes der gespielte Grundton, der durch eine Note einer Notenschrift vorgegeben ist, extrahiert.

[0021] Das erfindungsgemäße Konzept zeichnet sich somit dadurch aus, daß die Melodiesequenz, d. h. das Musiksignal, von einem beliebigen Musikinstrument vorgetragen werden kann. Das erfindungsgemäße Konzept ist robust gegenüber verstimmten Instrumenten, "schiefen" Tonlagen beim Singen oder Pfeifen von ungeübten Sängern und unterschiedlich vorgetragenen Tempi im zu bearbeitenden Liedausschnitt.

[0022] Ferner kann das Verfahren in seiner bevorzugten Ausführungsform, bei der eine Hough-Transforma-

tion zur Erzeugung der Frequenz-Zeit-Darstellung des Musiksignals eingesetzt wird, Rechenzeit-effizient implementiert werden, wodurch eine hohe Ausführungsgeschwindigkeit erreicht werden kann.

[0023] Ein weiterer Vorteil des erfindungsgemäßen Konzepts besteht darin, daß zur Referenzierung eines gesungenen oder gespielten Musiksignals aufgrund der Tatsache, daß eine Noten-basierte Beschreibung, die eine Rhythmus-Darstellung und eine Darstellung der Notenhöhen liefert, eine Referenzierung in einer Datenbank vorgenommen werden kann, in der eine Vielzahl von Musiksignalen abgespeichert sind. Insbesondere aufgrund der großen Verbreitung des MIDI-Standards existiert ein reicher Schatz an MIDI-Dateien für eine große Anzahl von Musikstücken.

[0024] Ein weiterer Vorteil des erfindungsgemäßen Konzepts besteht darin, daß auf der Basis der erzeugten Noten-basierten Beschreibung mit den Methoden der DNA-Sequenzierung Musikdatenbanken beispielsweise im MIDI-Format mit leistungskräftigen DNA-Sequenzierungs-Algorithmen, wie z. B. dem Boyer-Moore-Algorithmus, unter Verwendung von Replace/Insert/Delete-Operationen durchsucht werden können. Diese Form des zeitlich sequentiell ablaufenden Vergleichs unter gleichzeitiger gesteuerter Manipulation des Musiksignals liefert ferner die benötigte Robustheit gegenüber ungenauen Musiksignalen, wie sie durch ungeübte Instrumentalisten oder ungeübte Sänger erzeugt werden können. Dieser Punkt ist wesentlich für einen hohen Verbreitungsgrad eines Musikerkennungssystems, da die Anzahl geübter Instrumentalisten und geübter Sänger unter der Bevölkerung naturgemäß eher gering ist.

[0025] Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend bezugnehmend auf die beiliegenden Zeichnungen näher erläutert. Es zeigen:

Fig. 1 ein Blockschaltbild einer erfindungsgemäßen Vorrichtung zum Überführen eines Musiksignals in eine Noten-basierte Darstellung;

Fig. 2 ein Blockschaltbild einer bevorzugten Vorrichtung zum Erzeugen einer Frequenz-Zeit-Darstellung aus einem Musiksignal, bei der zur Flankendetektion eine Hough-Transformation eingesetzt wird;

Fig. 3 ein Blockschaltbild einer bevorzugten Vorrichtung zum Erzeugen einer segmentierten Zeit-Frequenz-Darstellung aus der durch Fig. 2 gelieferten Frequenz-Zeit-Darstellung;

Fig. 4 eine erfindungsgemäße Vorrichtung zum Ermitteln einer Folge von Notenhöhen auf der Basis der von Fig. 3 ermittelten segmentierten Zeit-Frequenz-Darstellung;

Fig. 5 eine bevorzugte Vorrichtung zum Ermitteln eines Noten-Rhythmus auf der Basis der segmentierten Zeit-Frequenz-Darstellung von Fig. 3;

Fig. 6 eine schematische Darstellung einer Design-Rule-Überprüfungseinrichtung, um unter Kenntnis der Notenhöhen und des Notenrhythmus zu überprüfen, ob die ermittelten Werte nach kompositorischen Regeln sinnvoll sind;

Fig. 7 ein Blockschaltbild einer erfindungsgemäßen Vorrichtung zum Referenzieren eines Musiksignal in einer Datenbank; und

Fig. 8 ein Frequenz-Zeit-Diagramm der ersten 13 Sekunden des Klarinettenquintetts A-Dur von W. A. Mozart, KV 581, Larghetto, Jack Bryner, Klarinette, Aufnahme: 12/1969, London, Philips 420 710-2 einschließlich Fitfunktion und Notenhöhen.

[0026] Fig. 1 zeigt ein Blockschaltbild einer erfindungsgemäßen Vorrichtung zum Überführen eines Musiksignals in eine Noten-basierte Darstellung. Ein Musiksignal, das gesungen, gespielt oder in Form von digitalen zeitlichen Abtastwerten vorliegt, wird in eine Einrichtung 10 zum Erzeugen einer Frequenz-Zeit-Darstellung des Musiksignals eingespeist, wobei die Frequenz-Zeit-Darstellung Koordinatentupel aufweist, wobei ein Koordinatentupel einen Frequenzwert und einen Zeitwert umfaßt, wobei der Zeitwert die Zeit des Auftretens der zugeordneten Frequenz in dem Musiksignal angibt. Die Frequenz-Zeit-Darstellung wird in eine Einrichtung 12 zum Berechnen einer Fitfunktion als Funktion der Zeit eingespeist, deren Verlauf durch die Koordinatentupel der Frequenz-Zeit-Darstellung bestimmt ist. Aus der Fitfunktion werden mittels einer Einrichtung 14 benachbarte Extrema ermittelt, welche dann von einer Einrichtung 16 zum Segmentieren der Frequenz-Zeit-Darstellung verwendet werden, um eine Segmentierung durchzuführen, die auf einen Notenrhythmus hinweist, der an einem Ausgang 18 ausgegeben wird. Die Segmentierungsinformationen werden ferner von einer Einrichtung 20 verwendet, die zur Bestimmung der Tonhöhe pro Segment vorgesehen ist. Die Einrichtung 20 verwendet zur Bestimmung der Tonhöhe pro Segment lediglich die Koordinaten-Tupel in einem Segment, um für die aufeinanderfolgenden Segmente aufeinanderfolgende Notenhöhen an einem Ausgang 22 auszugeben. Die Daten am Ausgang 18, also die Rhythmusinformationen, und die Daten an dem Ausgang 22, also die Ton- bzw. Notenhöheninformationen, bilden zusammen eine Noten-basierte Darstellung, aus der eine MIDI-Datei oder mittels einer graphischen Schnittstelle auch eine Notenschrift erzeugt werden kann.

[0027] Im nachfolgenden wird anhand von Fig. 2 auf eine bevorzugte Ausführungsform zum Erzeugen einer

Frequenz-Zeit-Darstellung des Musiksignals eingegangen. Ein Musiksignal, das beispielsweise als Folge von PCM-Samples vorliegt, wie sie durch Aufzeichnen eines gesungenen oder gespielten Musiksignals und anschließendes Abtasten und Analog/Digital-Wandeln erzeugt werden, wird in einen Audio-I/O-Handler 10a eingespeist. Alternativ kann das Musiksignal in digitalem Format auch direkt von der Festplatte eines Computers oder von der Soundkarte eines Computers kommen. Sobald der Audio-I/O-Handler 10a eine Ende-Datei-Marke erkennt, schließt er die Audiodatei und lädt je nach Bedarf das nächste zu bearbeitende Audiofile oder terminiert den Einlesevorgang. Die stromförmig vorliegenden PCM-Samples (PCM = Pulse Code Modulation) werden nacheinander an eine Vorverarbeitungseinrichtung 10b übermittelt, in der der Datenstrom auf eine einheitliche Abtastrate umgewandelt wird. Es wird bevorzugt, in der Lage zu sein, mehrere Abtastraten zu verarbeiten, wobei die Abtastrate des Signals bekannt sein soll, um aus der Abtastrate Parameter für die nachfolgende Signalflankendetektionseinheit 10c zu ermitteln.

[0028] Die Vorverarbeitungseinrichtung 10b umfaßt ferner eine Pegelanpassungseinheit, die allgemein eine Normierung der Lautstärke des Musiksignals durchführt, da die Lautstärkeinformation des Musiksignals in der Frequenz-Zeit-Darstellung nicht benötigt wird. Damit die Lautstärkeinformationen die Bestimmung der Frequenz-Zeit-Koordinatentupel nicht beeinflussen, wird eine Lautstärkenormierung folgendermaßen vorgenommen. Die Vorverarbeitungseinheit zur Normierung des Pegels des Musiksignals umfaßt einen Look-Ahead-Buffer und bestimmt daraus die mittlere Lautstärke des Signals. Das Signal wird dann mit einem Skalierungsfaktor multipliziert. Der Skalierungsfaktor ist das Produkt aus einem Gewichtungsfaktor und dem Quotienten aus Vollausschlag und mittlerer Signallautstärke. Die Länge des Look-Ahead-Buffers ist variabel.

[0029] Die Flankendetektionseinrichtung 10c ist angeordnet, um aus dem Musiksignal Signalflanken spezifizierter Länge zu extrahieren. Die Einrichtung 10c führt vorzugsweise eine Hough-Transformation durch.

[0030] Die Hough-Transformation ist in dem U.S.-Patent Nr. 3,069,654 von Paul V. C. Hough beschrieben. Die Hough-Transformation dient zur Erkennung von komplexen Strukturen und insbesondere zur automatischen Erkennung von komplexen Linien in Photographien oder anderen Bilddarstellungen. In ihrer Anwendung gemäß der vorliegenden Erfindung wird die Hough-Transformation dazu verwendet, um aus dem Zeitsignal Signalflanken mit spezifizierten zeitlichen Längen zu extrahieren. Eine Signalflanke wird zunächst durch ihre zeitliche Länge spezifiziert. Im Idealfall einer Sinuswelle wäre eine Signalflanke durch die ansteigende Flanke der Sinusfunktion von 0 bis 90° definiert. Alternativ könnte die Signalflanke auch durch den Anstieg der Sinus-Funktion von - 90° bis +90° spezifiziert sein.

[0031] Liegt das Zeitsignal als Folge von zeitlichen Abtastwerten vor, so entspricht die zeitliche Länge einer Signalflanke unter Berücksichtigung der Abtastfrequenz, mit der die Samples erzeugt worden sind, einer bestimmten Anzahl von Abtastwerten. Die Länge einer Signalflanke kann somit ohne weiteres durch Angabe der Anzahl der Abtastwerte, die die Signalflanke umfassen soll, spezifiziert werden.

[0032] Darüber hinaus wird es bevorzugt, eine Signalflanke nur dann als Signalflanke zu detektieren, wenn dieselbe stetig ist und einen monotonen Verlauf hat, also im Falle einer positiven Signalflanke einen monoton steigenden Verlauf hat. Selbstverständlich können auch negative Signalflanken, also monoton fallende Signalflanken detektiert werden.

[0033] Ein weiteres Kriterium zur Klassifizierung von Signalflanken besteht darin, daß eine Signalflanke nur dann als Signalflanke detektiert wird, wenn sie einen bestimmten Pegelbereich überstreicht. Um Rauschstörungen auszublenden, wird es bevorzugt, für eine Signalflanke einen minimalen Pegelbereich oder Amplitudenbereich vorzugeben, wobei monoton steigende Signalflanken unterhalb dieses Bereichs nicht als Signalflanken detektiert werden.

[0034] Die Signalflankendetektionseinheit 12 liefert somit eine Signalflanke und den Zeitpunkt des Auftretens der Signalflanke. Hierbei ist es unerheblich, ob als Zeitpunkt der Signalflanke der Zeitpunkt des ersten Abtastwerts der Signalflanke, der Zeitpunkt des letzten Abtastwerts der Signalflanke oder der Zeitpunkt irgendeines Abtastwerts innerhalb der Signalflanke genommen wird, so lange aufeinanderfolgende Signalflanken gleich behandelt werden.

[0035] Dem Flankendetektor 10c ist eine Frequenzberechnungseinheit 10d nachgeschaltet. Die Frequenzberechnungseinheit 10d ist ausgebildet, um zwei zeitlich aufeinander folgende gleiche oder innerhalb eines Toleranzwerts gleiche Signalflanken zu suchen und dann die Differenz der Auftrittszeiten der Signalflanken zu bilden. Der Kehrwert der Differenz entspricht der Frequenz, die durch die beiden Signalflanken bestimmt ist. Wenn ein einfacher Sinuston betrachtet wird, so ist eine Periode des Sinustons durch den zeitlichen Abstand zwei aufeinanderfolgender gleich langer z. B. positiver Signalflanken gegeben.

[0036] Es sei darauf hingewiesen, daß die Hough-Transformation eine hohe Auflösung beim Detektieren von Signalflanken in dem Musiksignal aufweist, so daß durch die Frequenzberechnungseinheit 10d eine Frequenz-Zeit-Darstellung des Musiksignals erhalten werden kann, die mit hoher Auflösung die zu einem bestimmten Zeitpunkt vorhandenen Frequenzen aufweist. Eine solche Frequenz-Zeit-Darstellung ist in Fig. 8 gezeigt. Die Frequenz-Zeit-Darstellung hat als Abszisse eine Zeitachse, entlang der die absolute Zeit in Sekunden aufgetragen ist, und hat als Ordinate eine Frequenzachse, in der bei der in Fig. 8 gewählten Darstellung die Frequenz in Hz aufgetragen ist. Sämtliche Bildpunkte in Fig. 8 stellen Zeit-Frequenz-Koordinatentupel dar, wie sie erhalten werden, wenn die ersten 13 Sekun-

den des Werks von W. A. Mozart, Köchel-Verzeichnis Nr. 581, einer Hough-Transformation unterzogen werden. In den ersten etwa 5,5 Sekunden dieses Stückes findet sich ein relativ polyphoner Orchesterpart mit einer großen Bandbreite von relativ gleichmäßig auftretenden Frequenzen zwischen etwa 600 und etwa 950 Hz. Dann, etwa ab 5,5 Sekunden, setzt eine dominante Klarinettenstimme ein, die die Tonfolge H1, C2, Cis2, D2, H1 und A1 spielt. Die Orchestermusik tritt gegenüber der Klarinette in den Hintergrund, was sich in der Frequenz-Zeit-Darstellung von Fig. 8 dadurch bemerkbar macht, daß die hauptsächliche Verteilung von Frequenz-Zeit-Koordinatentupeln innerhalb eines begrenzten Bandes 800 liegt, das auch als Pitch-Contour-Streifenband bezeichnet wird. Eine Häufung von Koordinatentupeln um einen Frequenzwert deutet darauf hin, daß das Musiksignal einen relativ monophonen Anteil hat, wobei zu beachten ist, daß übliche Blech/Holzblasinstrumente neben dem Grundton eine Vielzahl von Obertönen erzeugen, wie z. B. die Oktave, die nächste Quint, etc. Auch diese Obertöne werden mittels der Hough-Transformation und anschließender Frequenzberechnung durch die Einheit 10d ermittelt und tragen zu dem verbreiterten Pitch-Contour-Streifenband bei. Auch das Vibrato eines Musikinstruments, das sich durch eine schnelle Frequenzänderung über der Zeit des gespielten Tons auszeichnet, trägt zu einer Verbreiterung des Pitch-Contour-Streifenbands bei. Wird eine Folge von Sinustönen erzeugt, so würde das Pitch-Contour-Streifenband zu einer Pitch-Contour-Linie degenerieren.

[0037] Der Frequenzberechnungseinheit 10d ist eine Einrichtung 10e zur Ermittlung von Häufungsgebieten nachgeschaltet. In der Einrichtung 10e zur Ermittlung der Häufungsgebiete werden die charakteristischen Verteilungspunktwolken (Cluster), die sich bei der Bearbeitung von Audiodateien als stationäres Merkmal ergeben, herausgearbeitet. Hierzu kann eine Elimination aller isolierten Frequenz-Zeit-Tupel durchgeführt werden, welche einen vorgegebenen Mindestabstand zum nächsten räumlichen Nachbarn überschreiten. So wird eine solche Verarbeitung dazu führen, daß nahezu sämtliche Koordinatentupel oberhalb des Pitch-Contour-Streifenbands 800 eliminiert werden, wodurch am Beispiel von Fig. 8 in dem Bereich von 6 bis 12 Sekunden lediglich das Pitch-Contour-Streifenband und einige Häufungsgebiete unterhalb des Pitch-Contour-Streifenbands verbleiben.

[0038] Das Pitch-Contour-Streifenband 800 besteht somit aus Clustern bestimmter Frequenzbreite und zeitlicher Länge, wobei diese Cluster von den gespielten Tönen hervorgerufen werden.

[0039] Die durch die Einrichtung 10e erzeugte Frequenz-Zeit-Darstellung, in der die isolierten Koordinatentupel bereits eliminiert sind, wird vorzugsweise zur Weiterverarbeitung anhand der Vorrichtung, die in Fig. 3 gezeigt ist, verwendet. Alternativ könnte jedoch auf die Elimination von Tupeln außerhalb des Pitch-Contour-Streifenbands verzichtet werden, um ein Segmentieren der Zeit-Frequenz-Darstellung zu erreichen. Dies könnte jedoch dazu führen, daß die zu berechnende Fitfunktion "irre geführt" wird, und Extremwerte liefert, die nicht Tongrenzen zugeordnet sind, sondern die aufgrund der außerhalb des Pitch-Contour-Streifenbands liegende Koordinatentupel vorhanden sind.

[0040] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird, wie es in Fig. 3 gezeigt ist, eine instrumentenspezifische Nachbearbeitung 10f durchgeführt, um aus dem Pitch-Contour-Streifenband 800 wenn möglich eine einzige Pitch-Contour-Linie zu erzeugen. Hierzu wird das Pitch-Contour-Streifenband einer instrumentenspezifischen Fallanalyse unterzogen. Bestimmte Instrumente, wie z. B. Oboe oder Waldhorn, weisen charakteristische Pitch-Contour-Streifenbänder auf. Bei der Oboe treten beispielsweise zwei parallele Streifenbänder auf, da durch das Doppelrohrblatt des Oboen-Mundstücks die Luftsäule zu zwei Longitudinalschwingungen unterschiedlicher Frequenz angeregt wird, und die Schwingungsform zwischen diesen beiden Modi oszilliert. Die Einrichtung 10f zur instrumentenspezifischen Nachbearbeitung untersucht die Frequenz-Zeit-Darstellung auf das Vorliegen charakteristischer Merkmale, und schaltet, wenn diese Merkmale festgestellt wurden, ein instrumentenspezifisches Nachbehandlungsverfahren ein, welches auf beispielsweise in einer Datenbank gespeicherte Spezialitäten verschiedener Instrumente eingeht. Eine Möglichkeit würde beispielsweise sein, von den zwei parallelen Streifenbändern der Oboe entweder das obere oder das untere zu nehmen, oder, je nach Bedarf, einen Mittelwert oder Medianwert zwischen beiden Streifenbändern der weiteren Verarbeitung zugrunde zu legen. Prinzipiell ist es möglich, für einzelne Instrumente eigene Charakteristika im Frequenz-Zeit-Diagramm festzustellen, da jedes Instrument eine typische Klangfarbe aufweist, die durch die Zusammensetzung der Oberwellen und dem zeitlichen Verlauf der Grundfrequenz und der Oberwellen bestimmt ist.

[0041] Idealerweise wird am Ausgang der Einrichtung 10f eine Pitch-Contour-Linie, also ein sehr schmales Pitch-Contour-Streifenband erhalten. Im Falle eines polyphonen Klanggemisches mit dominanter monophoner Stimme, wie z. B. der Klarinettenstimme in der rechten Hälfte von Fig. 8, wird jedoch trotz instrumentenspezifischer Nachverarbeitung keine Pitch-Contour-Linie erreichbar sein, da auch die Hintergrundinstrumente Töne spielen, die zu einer Verbreiterung führen.

[0042] Im Falle einer monophonen Singstimme oder eines einzelnen Instruments ohne Hintergrundorchester liegt jedoch nach der instrumentenspezifischen Nachbearbeitung durch die Einrichtung 10f eine schmale Pitch-Contour-Linie vor.

[0043] An dieser Stelle sei darauf hingewiesen, daß die Frequenz-Zeit-Darstellung, wie sie beispielsweise hinter der Einheit 10d von Fig. 2 vorliegt, alternativ auch durch ein Frequenztransformationsverfahren erzeugt

werden kann, wie es beispielsweise eine schnelle Fourier-Transformation ist. Durch eine Fourier-Transformation wird aus einem Block von zeitlichen Abtastwertes des Musiksignals ein Kurzzeitspektrum erzeugt. Problematisch bei der Fourier-Transformation ist jedoch die Tatsache der geringen Zeitauflösung, wenn ein Block mit vielen Abtastwerten in den Frequenzbereich transformiert wird. Ein Block mit vielen Abtastwerten ist jedoch erforderlich, um eine gute Frequenzauflösung zu erreichen. Wird dagegen, um eine hohe Zeitauflösung zu erreichen, ein Block mit wenigen Abtastwerten verwendet, so wird eine geringere Frequenzauflösung erreicht. Daraus wird ersichtlich, daß bei einer Fourier-Transformation entweder eine hohe Frequenzauflösung oder eine hohe Zeitauflösung erreicht werden kann. Eine hohe Frequenz- und eine hohe Zeitauflösung schließen sich, wenn die Fourier-Transformation verwendet wird, gegenseitig aus. Wenn dagegen eine Flankendetektion mittels der Hough-Transformation und eine Frequenzberechnung, um die Frequenz-Zeit-Darstellung zu erhalten, durchgeführt wird, ist sowohl eine hohe Frequenzauflösung als auch eine hohe Zeitauflösung zu erreichen. Um einen Frequenzwert bestimmen zu können, benötigt die Vorgehensweise mit der Hough-Transformation lediglich z. B. zwei ansteigende Signalflanken und daher lediglich zwei Periodendauern. Im Gegensatz zur Fourier-Transformation wird die Frequenz jedoch mit hoher Auflösung bestimmt, wobei gleichzeitig eine hohe Zeitauflösung erreicht wird. Aus diesem Grund wird die Hough-Transformation zur Erzeugen der Frequenz-Zeit-Darstellung gegenüber einer Fourier-Transformation bevorzugt.

[0044] Um einerseits die Tonhöhe eines Tons zu bestimmen, und um andererseits den Rhythmus eines Musiksignals ermitteln zu können, muß aus der Pitch-Contour-Linie bestimmt werden, wann ein Ton beginnt und wann derselbe endet. Hierzu wird erfindungsgemäß eine Fitfunktion verwendet, wobei bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung eine Polynomfitfunktion mit einem Grad n verwendet wird.

[0045] Obgleich andere Fitfunktionen auf der Basis von beispielsweise Sinusfunktionen oder Exponentialfunktionen möglich sind, wird gemäß der vorliegenden Erfindung eine Polynomfitfunktion mit einem Grad n bevorzugt. Wenn eine Polynomfitfunktion verwendet wird, geben die Abstände zwischen zwei Minima der Polynomfitfunktion einen Hinweis auf die zeitliche Segmentierung des Musiksignals, d. h. auf die Folge von Noten des Musiksignals. Eine solche Polynomfitfunktion 820 ist in Fig. 8 eingezeichnet. Es ist zu sehen, daß die Polynomfitfunktion 820 zu Anfang des Musiksignals und nach etwa 2,8 Sekunden zwei Polynomfitnullstellen 830, 832 aufweist, welche die beiden polyphonen Häufungsgebiete am Beginn des Mozart-Stücks "einleiten". Dann geht das Mozart-Stück in eine monophone Gestalt über, da die Klarinette dominant gegenüber den begleitenden Streichern hervortritt und die Tonfolge h1 (Achtel), c2 (Achtel), cis2 (Achtel), d2 (punktierte Achtel), h1 (Sechzehntel) und a1 (Viertel) spielt. Entlang der Zeitachse sind die Minima der Polynomfitfunktion durch die kleinen Pfeile (z. B. 834) markiert. Obgleich es bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung bevorzugt wird, nicht unmittelbar das zeitliche Auftreten der Minima zur Segmentierung zu verwenden, sondern noch eine Skalierung mit einer vorher berechneten Skalierungskennlinie durchzuführen, führt auch bereits eine Segmentierung ohne Verwendung der Skalierungskennlinie zu brauchbaren Ergebnissen, wie es aus Fig. 8 zu sehen ist.

[0046] Die Koeffizienten der Polynomfitfunktion, welche einen hohen Grad im Bereich von über 30 aufweisen kann, werden mit Methoden der Ausgleichsrechnung unter Verwendung der Frequenz-Zeit-Koordinatentupel, die in Fig. 8 gezeigt sind, berechnet. Bei dem in Fig. 8 gezeigten Beispiel werden hierzu sämtliche Koordinatentupel verwendet. Die Polynomfitfunktion wird so in die Frequenz-Zeit-Darstellung gelegt, daß die Polynomfitfunktion in einem bestimmten Abschnitt des Stücks, in Fig. 8 die ersten 13 Sekunden, optimal in die Koordinaten-Tupel gelegt wird, so daß der Abstand der Tupel zur Polynomfitfunktion insgesamt gerechnet minimal wird. Dadurch können "Scheinminima" entstehen, wie beispielsweise das Minima der Polynomfitfunktion bei etwa 10,6 Sekunden. Dieses Minima rührt daher, daß unter dem Pitch-Contour-Streifenband Cluster sind, die bevorzugterweise durch die Einrichtung 10e zur Ermittlung der Häufungsgebiete (Fig. 2) beseitigt werden.

[0047] Nachdem die Koeffizienten der Polynomfitfunktion berechnet worden sind, können mittels einer Einrichtung 10h die Minima der Polynomfitfunktion bestimmt werden. Da die Polynomfitfunktion analytisch vorliegt, ist eine einfache Differenzierung und Nullstellensuche ohne weiteres möglich. Für andere Polynomfitfunktionen können numerische Verfahren zum Ableiten und Nullstellensuchen eingesetzt werden.

[0048] Wie es bereits ausgeführt worden ist, wird durch die Einrichtung 16 eine Segmentierung der Zeit-Frequenz-Darstellung auf der Basis der ermittelten Minima vorgenommen.

[0049] Im nachfolgenden wird darauf eingegangen, wie der Grad der Polynomfitfunktion, deren Koeffizienten durch die Einrichtung 12 berechnet werden, gemäß einem bevorzugten Ausführungsbeispiel bestimmt wird. Hierzu wird eine Standardtonfolge mit festgelegten Standardlängen zur Kalibrierung der erfindungsgemäßen Vorrichtung vorgespielt. Daraufhin wird für Polynome verschiedener Grade eine Koeffizientenberechnung und Minimaermittlung durchgeführt. Der Grad wird dann so gewählt, daß die Summe der Differenzen zweier aufeinanderfolgender Minima des Polynoms von der gemessenen Tonlänge, d. h. durch Segmentierung bestimmten Tonlänge, der vorgespielten Standardreferenztöne minimiert wird. Ein zu geringer Grad des Polynoms führt dazu, daß das Polynom zu ringer Grad des Polynoms führt dazu, daß das Polynom zu grob vorgeht

und den einzelnen Tönen nicht folgen kann, während ein zu hoher Grad des Polynoms dazu führen kann, daß die Polynomfitfunktion zu stark "zappelt". Bei dem in Fig. 8 gezeigten Beispiel wurde ein Polynom fünfzigster Ordnung gewählt. Diese Polynomfitfunktion wird dann für einen nachfolgenden Betrieb zugrunde gelegt, so daß die Einrichtung zum Berechnen der Fitfunktion (12 in Fig. 1) vorzugsweise lediglich die Koeffizienten der Polynomfitfunktion und nicht zusätzlich den Grad der Polynomfitfunktion berechnen muß, um eine Rechenzeitersparnis zu erreichen.

[0050] Der Kalibrierungslauf unter Verwendung der Tonfolge aus Standardreferenztönen vorgegebener Länge kann ferner dazu verwendet werden, um eine Skalierungskennlinie zu ermitteln, die in die Einrichtung 16 zum Segmentieren eingespeist werden kann (30), um den zeitlichen Abstand der Minima der Polynomfitfunktion zu skalieren. Wie es aus Fig. 8 ersichtlich ist, liegt das Minima der Polynomfitfunktion nicht unmittelbar am Beginn des Haufens, der den Ton h1 darstellt, also nicht unmittelbar bei etwa 5,5 Sekunden, sondern etwa bei 5,8 Sekunden. Wenn eine Polynomfitfunktion höherer Ordnung gewählt wird, würde das Minima mehr zum Rand des Haufens hin bewegt werden. Dies würde jedoch unter Umständen dazu führen, daß die Polynomfitfunktion zu stark zappelt und zu viele Scheinminima erzeugt. Daher wird es bevorzugt, die Skalierungskennlinie zu erzeugen, die für jeden berechneten Minimaabstand einen Skalierungsfaktor bereit hält. Je nach Quantelung der vorgespielten Standardreferenztöne kann eine Skalierungskennlinie mit frei wählbarer Auflösung erzeugt werden. Es sei darauf hingewiesen, daß diese Kalibrierungs- bzw. Skalierungskennlinie lediglich einmal vor Inbetriebnahme der Vorrichtung erzeugt werden muß, um dann während eines Betriebs der Vorrichtung zum Überführen eines Musiksignals in eine Notenbasierte Beschreibung verwendet werden zu können.

[0051] Die zeitliche Segmentierung der Einrichtung 16 erfolgt somit durch den Polynomfit n-ter Ordnung, wobei der Grad vor Inbetriebnahme der Vorrichtung so gewählt wird, daß die Summe der Differenzen zweier aufeinanderfolgender Minima des Polynoms von den gemessenen Tonlängen von Standardreferenztönen minimiert wird. Aus der mittleren Abweichung wird die Skalierungskennlinie bestimmt, die den Bezug zwischen der mit dem erfindungsgemäßen Verfahren gemessenen Tonlänge und der tatsächlichen Tonlänge herstellt. Obgleich ohne Skalierung bereits brauchbare Ergebnisse erhalten werden, wie es Fig. 8 deutlich macht, kann durch die Skalierungskennlinie die Genauigkeit des Verfahrens noch verbessert werden.

[0052] Im nachfolgenden wird auf Fig. 4 Bezug genommen, um einen bevorzugten Aufbau der Einrichtung 20 zum Bestimmen der Tonhöhe pro Segment darzustellen. Die durch die Einrichtung 16 von Fig. 3 segmentierte Zeit-Frequenz-Darstellung wird in eine Einrichtung 20a eingespeist, um einen Mittelwert aller Frequenz-Tupel oder aber einen Medianwert aller Koordinatentupel pro Segment zu bilden. Die besten Ergebnisse ergeben sich, wenn lediglich die Koordinatentupel innerhalb der Pitch-Contour-Linie verwendet werden. In der Einrichtung 20a wird somit für jeden Cluster, dessen Intervallgrenzen durch die Einrichtung 16 zum Segmentieren (Fig. 3) bestimmt worden sind, ein Pitchwert, d. h. ein Tonhöhenwert, gebildet. Das Musiksignal liegt am Ausgang der Einrichtung 20a somit bereits als eine Folge von absoluten Pitchhöhen vor. Prinzipiell könnte diese Folge von absoluten Pitchhöhen bereits als Notenfolge bzw. Noten-basierte Darstellung verwendet werden.

[0053] Um jedoch eine robustere Notenberechnung zu erhalten, und um von der Stimmung der verschiedenen Instrumente etc. unabhängig zu werden, wird anhand der Folge von Pitchwerten am Ausgang der Einrichtung 20a die absolute Stimmung, die durch die Angabe der Frequenzverhältnisse zweier benachbarter Halbtonstufen und den Referenzkammerton spezifiziert ist, bestimmt. Hierzu wird aus den absoluten Pitchwerten der Tonfolge ein Tonkoordinatensystem durch die Einrichtung 20b berechnet. Sämtliche Töne des Musiksignals werden genommen, und es werden sämtliche Töne von den anderen Tönen jeweils subtrahiert, um möglichst sämtliche Halbtöne der Tonleiter, die dem Musiksignal zugrunde liegt, zu erhalten. Beispielsweise sind die Intervallkombinationspaare für eine Notenfolge der Länge im einzelnen: Note 1 minus Note 2, Note 1 minus Note 3, Note 1 minus Note 4, Note 1 minus Note 5, Note 2 minus Note 3, Note 2 minus Note 4, Note 2 minus Note 5, Note 3 minus Note 4, Note 3 minus Note 5, Note 4 minus Note 5.

[0054] Der Satz von Intervallwerten bildet ein Tonkoordinatensystem. Dieses wird nunmehr in eine Einrichtung 20c eingespeist, die eine Ausgleichsrechnung durchführt und das durch die Einrichtung 20b berechnete Tonkoordinatensystem mit Tonkoordinatensystemen vergleicht, die in einer Stimmungen-Datenbank 40 gespeichert sind. Die Stimmung kann gleichschwebend (Unterteilung einer Oktave in 12 gleich große Halbtonintervalle), enharmonisch, natürlich harmonisch, pythagoräisch, mitteltönig, nach Huygens, zwölfteilig mit natürlicher harmonischer Basis nach Kepler, Euler, Mattheson, Kirnberger I + II, Malcolm, mit modifizierten Quinten nach Silbermann, Werckmeister III, IV; V, VI, Neidhardt I, II, III sein. Ebenso kann die Stimmung instrumentenspezifisch sein, bedingt durch die Bauart des Instruments, d. h. beispielsweise durch die Anordnung der Klappen und Tasten etc. Die Einrichtung 20c bestimmt mittels der Methoden der Ausgleichsrechnung die absoluten Halbtonstufen, indem durch Variationsrechnung die Stimmung angenommen wird, die die Gesamtsumme der Residuen der Abstände der Halbtonstufen von den Pitchwerten minimiert. Die absoluten Tonstufen werden dadurch bestimmt, daß die Halbtonstufen parallel in Schritten von 1 Hz geändert werden und diejenigen Halbtonstufen als absolut angenommen werden, die die Gesamtsumme der Residuen der Ab-

stände der Halbtonstufen von den Pitchwerten minimieren. Für jeden Pitchwert ergibt sich dann ein Abweichungswert von der nächstliegenden Halbtonstufe. Extremausreißer sind dadurch bestimmbar, wobei diese Werte ausgeschlossen werden können, indem iterativ ohne die Ausreißer die Stimmung neu berechnet wird. Am Ausgang der Einrichtung 20c liegt somit für jeden Pitchwert eines Segments eine nächstliegende Halbtonstufe der dem Musiksignal zugrunde liegenden Stimmung vor. Durch eine Einrichtung 20d zum Quantisieren wird der Pitchwert durch die nächstliegende Halbtonstufe ersetzt, so daß am Ausgang der Einrichtung 20d eine Folge von Notenhöhen sowie Informationen über die Stimmung, die dem Musiksignal zugrunde liegt, und den Referenzkammerton vorliegen. Diese Informationen am Ausgang der Einrichtung 20c könnten nunmehr ohne weiteres dazu verwendet werden, um Notenschrift zu erzeugen, oder um eine MIDI-Datei zu schreiben.

[0055] Es sei darauf hingewiesen, daß die Quantisierungseinrichtung 20d bevorzugt wird, um unabhängig von dem Instrument, das das Musiksignal liefert, zu werden. Wie es nachfolgend anhand von Fig. 7 dargestellt werden wird, ist die Einrichtung 20d vorzugsweise ferner ausgestaltet, um nicht nur die absoluten quantisierten Pitchwerte auszugeben, sondern um auch die Intervallhalbtonsprünge zwei aufeinanderfolgender Noten zu bestimmen und diese Folge von Halbtonsprüngen dann als Suchfolge für einen bezugnehmend auf Fig. 7 beschriebenen DNA-Sequenzer zu verwenden. Da das vorgespielte oder vorgesungene Musiksignal in eine andere Tonart transponiert sein kann, abhängig auch von der Grundstimmung des Instruments (z. B. B-Klarinette, Es-Saxophon), wird für die bezugnehmend auf Fig. 7 beschriebene Referenzierung nicht die Folge von absoluten Tonhöhen verwendet, sondern die Folge von Differenzen, da die Differenzfrequenzen von der absoluten Tonhöhe unabhängig sind.

[0056] Im nachfolgenden wird anhand von Fig. 5 auf eine bevorzugte Ausgestaltung der Einrichtung 16 zum Segmentieren der Frequenz-Zeit-Darstellung Bezug genommen, um den Notenrhythmus zu erzeugen. So könnten zwar bereits die Segmentierungsinformationen als Rhythmusinformationen verwendet werden, da durch dieselben die Dauer eines Tons gegeben ist. Es wird jedoch bevorzugt, die segmentierte Zeit-Frequenz-Darstellung bzw. die aus derselben durch Abstand zwei benachbarter Minima bestimmten Tonlängen mittels einer Einrichtung 16a in normierte Tonlängen zu transformieren. Diese Normierung wird mittels einer Subjective-Duration-Kennlinie aus der Tonlänge berechnet. So zeigen psychoakustische Forschungen, daß beispielsweise eine 1/8-Pause länger als eine 1/8-Note dauert. Solche Informationen gehen in die Subjective-Duration-Kennlinie ein, um die normierten Tonlängen und damit auch die normierten Pausen zu erhalten. Die normierten Tonlängen werden dann in eine Einrichtung 16b zur Histogrammierung eingespeist. Die Einrichtung 16b liefert eine Statistik darüber, welche Tonlängen auftreten bzw. um welche Tonlängen Häufungen stattfinden. Auf der Basis des Tonlängenhistogramms wird durch eine Einrichtung 16c eine Grundnotenlänge festgelegt, indem die Unterteilung der Grundnotenlänge so vorgenommen wird, daß die Notenlängen als ganzzahlige Vielfache dieser Grundnotenlänge angebbar sind. So kann man zu Sechzehntel-, Achtel-, Viertel-, Halb- oder Vollnoten gelangen. Die Einrichtung 16c basiert darauf, daß in üblichen Musiksignalen keineswegs beliebige Tonlängen vorgegeben sind, sondern die verwendeten Notenlängen üblicherweise in einem festen Verhältnis zueinander stehen.

[0057] Nachdem die Grundnotenlänge festgelegt worden ist und damit auch die zeitliche Länge von Sechzehntel-, Achtel-, Viertel-, Halb- oder Vollnoten werden die durch die Einrichtung 16a berechneten normierten Tonlängen in einer Einrichtung 16d dahingehend quantisiert, daß jede normierte Tonlänge durch die nächstliegende durch die Grundnotenlänge bestimmte Tonlänge ersetzt wird. Damit liegt eine Folge von quantisierten normierten Tonlängen vor, welche vorzugsweise in einen Rhythmus-Fitter/Takt-Modul 16e eingespeist wird. Der Rhythmus-Fitter bestimmt die Taktart, indem er berechnet, ob mehrere Noten zusammengefaßt jeweils Gruppen von Dreiviertelnoten, Vierviertelnoten, etc. bilden. Als Taktart wird diejenige angenommen, bei der ein über die Anzahl der Noten normiertes Maximum an richtigen Einträgen vorliegt.

[0058] Damit liegen Notenhöheninformationen und Notenrhythmusinformationen an den Ausgängen 22 (Fig. 4) und 18 (Fig. 5) vor. Diese Informationen können in einer Einrichtung 60 zur Design-Rule-Überprüfung zusammengeführt werden. Die Einrichtung 60 überprüft, ob die gespielten Tonfolgen nach kompositorischen Regeln der Melodieführung aufgebaut sind. Noten in der Folge, die nicht in das Schema passen, werden markiert, damit diese markierten Noten von dem DNA-Sequenzer, der anhand von Fig. 7 dargestellt wird, gesondert behandelt werden. Die Einrichtung 16 sucht nach sinnvollen Konstrukten und ist ausgebildet, um beispielsweise zu erkennen, ob bestimmte Notenfolgen unspielbar sind bzw. üblicherweise nicht auftreten.

[0059] Im nachfolgenden wird auf Fig. 7 Bezug genommen, um ein Verfahren zum Referenzieren eines Musiksignals in einer Datenbank gemäß einem weiteren Aspekt der vorliegenden Erfindung darzustellen. Das Musiksignal liegt am Eingang beispielsweise als Datei 70 vor. Durch eine Einrichtung 72 zum Überführen des Musiksignals in eine Noten-basierte Beschreibung, die gemäß den Fig. 1 bis 6 erfindungsgemäß aufgebaut ist, werden Notenrhythmus-Informationen und/oder Notenhöhen-Informationen erzeugt, die eine Suchfolge 74 für einen DNA-Sequenzer 76 bilden. Die Folge von Noten, die durch die Suchfolge 74 dargestellt ist, wird nunmehr entweder hinsichtlich des Notenrhythmus und/oder hinsichtlich der Notenhöhen mit einer Vielzahl von Noten-basierten Beschreibungen für verschiedene Stücke (Track_1 bis Track_n) verglichen, die in einer

Notendatenbank 78 abgespeichert sein können. Der DNA-Sequenzer, der eine Einrichtung zum Vergleichen des Musiksignals mit einer Noten-basierten Beschreibung der Datenbank 78 darstellt, prüft eine Übereinstimmung bzw. Ähnlichkeit. Somit kann eine Aussage hinsichtlich des Musiksignals auf der Basis des Vergleichs getroffen werden. Der DNA-Sequenzer 76 ist vorzugsweise mit einer Musik-Datenbank verbunden, in der die verschiedenen Stücke (Track_1 bis Track_n), deren Noten-basierte Beschreibungen in der Notendatenbank gespeichert sind, als Audiodatei abgelegt sind. Selbstverständlich können die Notendatenbank 78 und die Datenbank 80 eine einzige Datenbank sein. Alternativ könnte auch auf die Datenbank 80 verzichtet werden, wenn der Notendatenbank Metainformationen über die Stücke, deren Noten-basierten Beschreibungen abgespeichert sind, umfassen, wie z. B. Autor, Name des Stücks, Musikverlag, Pressung, etc.

[0060] Allgemein wird durch die in Fig. 7 gezeigte Vorrichtung eine Referenzierung eines Lieds erreicht, bei dem ein Audiofileabschnitt, in dem eine gesungene oder mit einem Musikinstrument gespielte Tonfolge aufgezeichnet ist, in eine Folge von Noten überführt wird, wobei diese Folge von Noten als Suchkriterium mit gespeicherten Notenfolgen in der Notendatenbank verglichen wird und das Lied aus der Notendatenbank referenziert wird, bei dem die größte Übereinstimmung zwischen Noteneingabefolge und Notenfolge in der Datenbank vorliegt. Als Noten-basierte Beschreibung wird die MIDI-Beschreibung bevorzugt, da MIDI-Dateien für riesige Mengen von Musikstücken bereits existieren. Alternativ könnte die in Fig. 7 gezeigte Vorrichtung auch aufgebaut sein, um die Noten-basierte Beschreibung selbst zu erzeugen, wenn die Datenbank zunächst in einem Lern-Modus betrieben wird, der durch einen gestrichelten Pfeil 82 angedeutet ist. Im Lern-Modus (82) würde die Einrichtung 72 zunächst für eine Vielzahl von Musiksignalen eine Noten-basierte Beschreibung erzeugen und in der Notendatenbank 78 abspeichern. Erst wenn die Notendatenbank ausreichend gefüllt ist, würde die Verbindung 82 unterbrochen werden, um eine Referenzierung eines Musiksignals durchzuführen. Nachdem MIDI-Dateien bereits für viele Stücke vorliegen, wird es jedoch bevorzugt, auf bereits vorhandene Notendatenbanken zurückzugreifen.

[0061] Insbesondere sucht der DNA-Sequenzer 76 die ähnlichste Melodietonfolge in der Notendatenbank, indem er die Melodietonfolge durch die Operationen Replace/Insert/Delete variiert. Jede Elementaroperation ist mit einem Kostenmaß verbunden. Optimal ist, wenn alle Noten ohne spezielle Operationen übereinstimmen. Suboptimal ist es dagegen, wenn n von m Werte übereinstimmen. Dadurch wird gewissermaßen automatisch ein Ranking der Melodiefolgen eingeführt, und die Ähnlichkeit des Musiksignals 70 zu einem Datenbank-Musiksignal Track_1 ... Track_n kann quantitativ angegeben werden. Es wird bevorzugt, die Ähnlichkeit von beispielsweise den besten fünf Kandidaten aus der Notendatenbank als absteigende Liste auszugeben.

[0062] In der Rhythmusdatenbank werden die Noten als Sechzehntel-, Achtel-, Viertel-, Halb- und Vollton abgelegt. Der DNA-Sequenzer sucht die ähnlichste Rhythmusfolge in der Rhythmusdatenbank, indem er die Rhythmusfolge durch die Operationen Replace/Insert/Delete variiert. Jede Elementaroperation ist ebenfalls wieder mit einem Kostenmaß verbunden. Optimal ist, wenn alle Notenlängen übereinstimmen, suboptimal ist es, wenn n von m Werte übereinstimmen. Dadurch wird wieder ein Ranking der Rhythmusfolgen eingeführt, und die Ähnlichkeit der Rhythmusfolgen kann in einer absteigenden Liste ausgegeben werden.

[0063] Der DNA-Sequenzer umfaßt bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung ferner eine Melodie/Rhythmus-Abgleicheinheit, die feststellt, welche Folgen sowohl von der Pitchfolge als auch von der Rhythmusfolge zusammen passen. Die Melodie/Rhythmus-Abgleicheinheit sucht die größtmögliche Übereinstimmung beider Folgen, indem die Zahl der Matches als Referenzkriterium angenommen wird. Optimal ist es, wenn alle Werte übereinstimmen, suboptimal ist es, wenn n von m Werte übereinstimmen. Dadurch wird wieder ein Ranking eingeführt, und die Ähnlichkeit der Melodie/Rhythmusfolgen kann wieder in einer absteigenden Liste ausgegeben werden.

[0064] Der DNA-Sequenzer kann ferner angeordnet sein, um von dem Design-Rule-Checker 60 (Fig. 6) markierte Noten entweder zu ignorieren bzw. mit einer geringeren Gewichtung zu versehen, damit das Ergebnis nicht durch Ausreißer unnötig verfälscht wird.

Patentansprüche

1. Verfahren zum Überführen eines Musiksignals in eine Noten-basierte Beschreibung, mit folgenden Schritten:

   Erzeugen (10) einer Frequenz-Zeit-Darstellung des Musiksignals, wobei die Frequenz-Zeit-Darstellung Koordinatentupel aufweist, wobei ein Koordinatentupel einen Frequenzwert und einen Zeitwert umfaßt, wobei der Zeitwert die Zeit des Auftretens der zugeordneten Frequenz in dem Musiksignal angibt;

   Berechnen (12) einer Fitfunktion als Funktion der Zeit, deren Verlauf durch die Koordinatentupel der Frequenz-Zeit-Darstellung bestimmt ist;

   Ermitteln (14) zumindest zwei benachbarter Extrema der Fitfunktion;

   zeitliches Segmentieren (16) der Frequenz-Zeit-Darstellung auf der Basis der ermittelten Extrema, wobei ein Segment durch zwei be-

nachbarte Extrema der Fitfunktion begrenzt, wobei die zeitliche Länge des Segments auf eine zeitliche Länge einer diesem Segment zugeordneten Note hinweist; und

Bestimmen (20) einer Tonhöhe der Note für das Segment unter Verwendung von Koordinaten-Tupeln in dem Segment.

2. Verfahren nach Anspruch 1, bei dem die Fitfunktion eine analytische Funktion ist, wobei die Einrichtung (14) zum Ermitteln benachbarter Extrema eine Differenzierung der analytischen Funktion und Nullstellenbestimmung durchführt.

3. Verfahren nach Anspruch 1 oder 2, bei dem die Extremwerte, die durch die Einrichtung (14) ermittelt werden, Minima der Fitfunktion sind.

4. Verfahren nach einem der vorhergehenden Ansprüche, bei dem die Fitfunktion eine Polynomfitfunktion des Grads n ist, wobei n größer als 2 ist.

5. Verfahren nach einem der vorhergehenden Ansprüche, bei dem im Schritt des Segmentierens (16) die zeitliche Länge einer Note unter Verwendung eines Kalibrierwerts aus dem zeitlichen Abstand zweier benachbarter Extremwerte bestimmt wird, wobei der Kalibrierwert das Verhältnis einer vorgegebenen zeitlichen Länge eines Tons zu einem Abstand zwischen zwei Extremwerten, der für den Ton unter Verwendung der Fitfunktion bestimmt wurde, ist.

6. Verfahren nach Anspruch 4 oder 5, bei dem der Grad der Fitfunktion unter Verwendung von vorgegebenen Tönen verschiedener bekannter Längen und für Fitfunktionen verschiedener Grade im voraus bestimmt wird, wobei der Grad im Schritt des Berechnens (12) verwendet wird, für den sich eine spezifizierte Übereinstimmung zwischen durch benachbarte Extremwerte bestimmten Tonlängen und bekannten Tonlängen ergibt.

7. Verfahren nach einem der Ansprüche 3 bis 6, bei dem im Schritt des zeitlichen Segmentierens (16) nur an einem solchen Minima der Fitfunktion segmentiert wird, dessen Frequenzwert zu dem Frequenzwert eines benachbarten Maximas um mindestens einen Minima-Maxima-Schwellenwert unterschiedlich ist, um Schein-Minima zu eliminieren.

8. Verfahren nach einem der vorhergehenden Ansprüche, bei dem im Schritt des Erzeugens (10) folgende Schritte durchgeführt werden:

Detektieren (10c) des zeitlichen Auftretens von Signalflanken in dem Zeitsignal;

Ermitteln (10d) eines zeitlichen Abstands zwischen zwei ausgewählten detektierten Signalflanken und Berechnen eines Frequenzwerts aus dem ermittelten zeitlichen Abstand und Zuordnen des Frequenzwerts zu einer Auftrittszeit des Frequenzwerts in dem Musiksignal, um einen Koordinatentupel aus dem Frequenzwert und der Auftrittszeit für diesen Frequenzwert zu erhalten.

9. Verfahren nach Anspruch 8, bei dem im Schritt des Detektierens (10c) eine Hough-Transformation durchgeführt wird.

10. Verfahren nach einem der vorhergehenden Ansprüche, bei dem im Schritt des Erzeugens (10) die Frequenz-Zeit-Darstellung gefiltert wird (10e), so daß ein Pitch-Contour-Streifenband verbleibt, und bei dem im Schritt des Berechnens (12) einer Fitfunktion lediglich die Koordinatentupel in dem Pitch-Contour-Streifenband berücksichtigt werden.

11. Verfahren nach einem der vorhergehenden Ansprüche, bei dem das Musiksignal monophon oder polyphon mit dominantem monophonen Anteil ist.

12. Verfahren nach Anspruch 11, bei dem das Musiksignal eine gesungene oder eine mit einem Instrument gespielte Notenfolge ist.

13. Verfahren nach einem der vorhergehenden Ansprüche, bei dem im Schritt (10) des Erzeugens einer Frequenz-Zeit-Darstellung eine Abtastratenumwandlung auf eine vorbestimmte Abtastrate durchgeführt wird (10b).

14. Verfahren nach einem der vorhergehenden Ansprüche, bei dem im Schritt (10) des Erzeugens einer Frequenz-Zeit-Darstellung eine Lautstärkenormierung (10b) durch Multiplikation mit einem Skalierungsfaktor, der von der mittleren Lautstärke eines Abschnitts und einer vorbestimmten maximalen Lautstärke abhängt, durchgeführt wird.

15. Verfahren nach einem der vorhergehenden Ansprüche, bei dem im Schritt des Erzeugens (10) eine instrumentenspezifische Nachbehandlung (10f) der Frequenz-Zeit-Darstellung durchgeführt wird, um eine instrumentenspezifische Frequenz-Zeit-Darstellung zu erhalten, und
bei dem im Schritt des Berechnens (12) der Fitfunktion die instrumentenspezifische Frequenz-Zeit-Darstellung zugrunde gelegt wird.

16. Verfahren nach einem der vorhergehenden Ansprüche, bei dem im Schritt des Bestimmens (20) der Tonhöhe pro Segment der Mittelwert der Koordinatentupel in einem Segment oder der Medianwert der

Koordinatentupel in dem Segment verwendet wird, wobei der Mittelwert oder der Medianwert in einem Segment auf einen absoluten Tonhöhenwert der Note für das Segment hinweist.

17. Verfahren nach Anspruch 16, bei dem der Schritt des Bestimmens (20) der Tonhöhe den Schritt des Ermittelns (20b, 20c) einer dem Musiksignal zugrunde liegenden Stimmung unter Verwendung der absoluten Tonhöhenwerte von Noten für Segmente des Musiksignals aufweist.

18. Verfahren nach Anspruch 17, bei dem der Schritt des Ermittelns der Stimmung folgende Merkmale aufweist:

Bilden (20b) einer Mehrzahl von Frequenzdifferenzen aus den Tonhöhenwerten des Musiksignals, um ein Frequenzdifferenz-Koordinatensystem zu erhalten;
Ermitteln (20c) der absoluten Stimmung, die dem Musiksignal zugrunde liegt, unter Verwendung des Frequenzdifferenzkoordinatensystems und unter Verwendung einer Mehrzahl von abgespeicherten Stimmungskoordinatensystemen (40) mittels einer Ausgleichsrechnung.

19. Verfahren nach Anspruch 18, bei dem der Schritt des Bestimmens (20) der Tonhöhe einen Schritt des Quantisierens (20d) der absoluten Tonhöhenwerte auf der Basis der absoluten Stimmung und des Referenz-Kammertons aufweist, um eine Note pro Segment zu erhalten.

20. Verfahren nach einem der vorhergehenden Ansprüche, bei dem der Schritt des Segmentierens (16) folgenden Schritt aufweist:

Transformieren (16a) der zeitlichen Länge von Tönen in normierte Notenlängen durch Histogrammieren (16b) der zeitlichen Länge und Festlegen (16c) einer Grundnotenlänge, derart, daß die zeitlichen Längen der Töne als ganzzahlige Vielfache oder ganzzahlige Bruchteile der Grundnotenlänge angebbar sind, und Quantisieren (16c) der zeitlichen Längen der Töne auf das nächstliegende ganzzahlige Vielfache oder den nächstliegenden ganzzahligen Bruchteil, um quantisierte Notenlängen zu erhalten.

21. Verfahren nach Anspruch 20, bei dem der Schritt des Segmentierens (16) ferner einen Schritt des Bestimmens (16e) eines Takts aus den quantisierten Notenlängen durch Untersuchen umfaßt, ob aufeinanderfolgende Noten zu einem Taktschema gruppiert werden können.

22. Verfahren nach Anspruch 21, das ferner folgenden Schritt aufweist:

Untersuchen (60) einer Folge von Noten, die das Musiksignal darstellt, wobei jede Note durch Anfang, Länge und Tonhöhe spezifiziert ist, hinsichtlich kompositorischer Regeln und Markieren einer Note, die mit den kompositorischen Regeln nicht vereinbar ist.

23. Verfahren zum Referenzieren eines Musiksignals (70) in einer Datenbank (78), die eine Noten-basierte Beschreibung einer Mehrzahl von Datenbank-Musiksignalen aufweist, mit folgenden Schritten:

Überführen (72) des Musiksignals in eine Noten-basierte Beschreibung (74) gemäß einem der Patentansprüche 1 bis 22;

Vergleichen (76) der Noten-basierten Beschreibung (74) des Musiksignals mit der Noten-basierten Beschreibung der Mehrzahl von Datenbank-Musiksignalen in der Datenbank (78); und

Treffen (76) einer Aussage hinsichtlich des Musiksignals (70) auf der Basis des Schritts des Vergleichens.

24. Verfahren nach Anspruch 23, bei dem die Noten-basierte Beschreibung für die Datenbank-Musiksignale ein MIDI-Format hat, wobei ein Tonanfang und ein Tonende als Funktion der Zeit spezifiziert sind, und bei dem vor dem Schritt des Vergleichens folgende Schritte ausgeführt werden:

Bilden von Differenzwerten zwischen zwei benachbarten Noten des Musiksignals, um eine Differenz-Notenfolge zu erhalten;

Bilden von Differenzwerten zwischen zwei benachbarten Noten der Noten-basierten Beschreibung des Datenbank-Musiksignals, und

bei dem im Schritt des Vergleichens die Differenz-Notenfolge des Musiksignals mit der Differenz-Notenfolge eines Datenbank-Musiksignals verglichen wird.

25. Verfahren nach Anspruch 23 oder 24, bei dem der Schritt des Vergleichens (76) unter Verwendung eines DNA-Sequenzing-Algorithmus und insbesondere unter Verwendung des Boyer-Moore-Algorithmus durchgeführt wird.

26. Verfahren nach einem der Ansprüche 23 bis 25, bei dem der Schritt des Treffens einer Aussage das Feststellen der Identität des Musiksignals (70) und

eines Datenbank-Musiksignals aufweist, falls die Noten-basierte Beschreibung des Datenbank-Musiksignals und die Noten-basierte Beschreibung des Musiksignals identisch sind.

27. Verfahren nach einem der Ansprüche 23 bis 25, bei dem der Schritt des Treffens einer Aussage hinsichtlich des Musiksignals eine Ähnlichkeit zwischen dem Musiksignal (70) und einem Datenbank-Musiksignal feststellt, wenn nicht alle Tonhöhen und/oder Tonlängen des Musiksignals mit Tonhöhen und/oder Tonlängen des Datenbank-Musiksignals übereinstimmen.

28. Verfahren nach einem der Ansprüche 23 bis 27, bei dem die Noten-basierte Beschreibung eine Rhythmusbeschreibung aufweist, und bei dem im Schritt des Vergleichens (76) ein Vergleich der Rhythmen des Musiksignals und des Datenbank-Musiksignals durchgeführt wird.

29. Verfahren nach einem der Ansprüche 23 bis 28, bei dem die Noten-basierte Beschreibung eine Tonhöhenbeschreibung aufweist, und bei dem im Schritt des Vergleichens (76) die Tonhöhen des Musiksignals mit den Tonhöhen eines Datenbank-Musiksignals verglichen werden.

30. Verfahren nach einem der Ansprüche 25 bis 29, bei dem im Schritt des Vergleichens (26) Einfügen-, Ersetzenoder Löschen-Operationen mit der Noten-basierten Beschreibung (74) des Musiksignals (70) durchgeführt werden, und bei dem im Schritt des Treffens einer Aussage eine Ähnlichkeit zwischen dem Musiksignal (70) und einem Datenbank-Musiksignal auf der Basis der Anzahl von Einfügen-, Ersetzen- oder Löschen-Operationen festgestellt wird, die erforderlich sind, um eine größtmögliche Übereinstimmung zwischen der Noten-basierten Beschreibung (74) des Musiksignals (70) und der Noten-basierten Beschreibung eines Datenbank-Musiksignals zu erreichen.

31. Vorrichtung zum Überführen eines Musiksignals in eine Noten-basierte Beschreibung, mit folgenden Merkmalen:

einer Einrichtung zum Erzeugen (10) einer Frequenz-Zeit-Darstellung des Musiksignals, wobei die Frequenz-Zeit-Darstellung Koordinatentupel aufweist, wobei ein Koordinatentupel einen Frequenzwert und einen Zeitwert umfaßt, wobei der Zeitwert die Zeit des Auftretens der zugeordneten Frequenz in dem Musiksignal angibt;

einer Einrichtung zum Berechnen (12) einer Fitfunktion als Funktion der Zeit, deren Verlauf durch die Koordinatentupel der Frequenz-Zeit-Darstellung bestimmt ist;

einer Einrichtung zum Ermitteln (14) zumindest zwei benachbarter Extrema der Fitfunktion;

einer Einrichtung zum zeitlichen Segmentieren (16) der Frequenz-Zeit-Darstellung auf der Basis der ermittelten Extrema, wobei ein Segment durch zwei benachbarte Extrema der Fitfunktion begrenzt, wobei die zeitliche Länge des Segments auf eine zeitliche Länge einer diesem Segment zugeordneten Note hinweist; und

einer Einrichtung zum Bestimmen (20) einer Tonhöhe der Note für das Segment unter Verwendung von Koordinaten-Tupeln in dem Segment.

32. Vorrichtung zum Referenzieren eines Musiksignal (70) in einer Datenbank (78), die eine Noten-basierte Beschreibung einer Mehrzahl von Datenbank-Musiksignalen aufweist, mit folgenden Merkmalen:

einer Einrichtung zum Überführen (72) des Musiksignals in eine Noten-basierte Beschreibung (74) durch ein Verfahren gemäß einem der Patentansprüche 1 bis 22;

einer Einrichtung zum Vergleichen (76) der Noten-basierten Beschreibung (74) des Musiksignals mit der Noten-basierten Beschreibung der Mehrzahl von Datenbank-Musiksignalen in der Datenbank (78); und

einer Einrichtung zum Treffen (76) einer Aussage hinsichtlich des Musiksignals (70) auf der Basis des Schritts des Vergleichens.

Claims

1. Method for transferring a music signal into a note-based description, comprising the following steps:

generating (10) a frequency-time representation of the music signal, with the frequency-time representation comprising coordinate tuples, with one coordinate tuple including a frequency value and a time value, with the time value indicating the time of occurrence of the assigned frequency in the music signal;

calculating (12) of a fit function as a function of time, the course of which is determined by the coordinate tuples of the frequency-time representation;

determining (14) of at least two adjacent extreme values of the fit function;

time-segmenting (16) of the frequency-time representation on the basis of the determined extreme values, with a segment being limited by two adjacent extreme values of the fit function, with the time length of the segment indicating a time length of a note assigned to this segment; and

determining (20) a tone height of the note for the segment using coordinate tuples in the segment.

2. Method in accordance with claim 1, wherein the fit function is an analytical function, with the means (14) for determining adjacent extreme values carrying out a deviation of the analytical function and a zero determination.

3. Method in accordance with claim 1 or 2, wherein the extreme values, which are determined by the means (14), are minimum values of the fit function.

4. Method in accordance with one of the preceding claims, in which the fit function is a polynomial fit function of the degree n, with n being greater than 2.

5. Method in accordance with one of the preceding claims, wherein, during the step of segmenting (16), the time length of a note is determined using a calibrating value from the time distance of two adjacent extreme values, with the calibrating value being the relationship of a specified time length of a tone to a distance between two extreme values, which was determined for the tone using the fit function.

6. Method in accordance with claim 4 or 5, in which the degree of the fit function using specified tones of varying known lengths and for fit functions of varying degrees is determined in advance, with the degree in the step of calculating (12) being used, for which a specified matching between tone lengths determined by adjacent extreme values and known tone lengths results.

7. Method in accordance with one of claims 3 to 6, wherein in the step of time-segmenting (16) only one such minimum value of the fit function is segmented, the frequency value of which is different from the frequency value of an adjacent maximum value by at least one minimum-maximum threshold value to eliminate fake minimum values.

8. Method in accordance with one of the preceding claims, wherein in the step of generating (10) the following steps are carried out:

detecting (10c) of the time occurrence of signal edges in the time signal;

determining (10d) a time distance between two selected detected signal edges and calculating a frequency value from the determined time distance and assigning the frequency value to an occurrence time of the frequency value in the music signal to obtain a coordinate tuple from the frequency value and the occurrence time for this frequency value.

9. Method in accordance with claim 8, wherein, in the step of detecting (10c), a Hough transform is carried out.

10. Method in accordance with one of the preceding claims, wherein, in the step of generating (10), the frequency-time representation is filtered (10e) such that a pitch-contour strip band remains, and, wherein, in the step of calculating (12) of a fit function, only the coordinate tuples in the pitch-contour strip band are considered.

11. Method in accordance with one of the preceding claims, wherein the music signal is monophonic or polyphonic with a dominant monophonic share.

12. Method in accordance with claim 11, wherein the music signal is a note sequence sung by a person or performed by an instrument.

13. Method in accordance with one of the preceding claims, wherein, in the step (10) of generating a frequency-time representation, a sample rate conversion is carried out to a predetermined sampled rate (10b).

14. Method in accordance with one of the preceding claims, wherein, in the step (10) of generating a frequency-time representation, a sound volume standardization (10b) is carried out by multiplication with a scaling factor, the scaling factor depending on a median sound volume of a section or a predetermined maximum sound volume.

15. Method in accordance with one of the preceding claims, wherein, in the step of generating (10), an instrument-specific postprocessing (10f) of the frequency-time representation is carried out to obtain an instrument-specific frequency-time representation, and
wherein, the step of calculating (12) of the fit function is based on the instrument-specific frequency-time representation.

16. Method in accordance with one of the preceding claims, wherein, in the step of determining (20) the

tone height per segment, the mean value of the coordinate tuple in a segment or the median value of the coordinate tuple in the segment is used, with the mean value or the median value in a segment indicating an absolute tone height value of the note for the segment.

17. Method in accordance with claim 16, wherein the step of determining (20) the tone height comprises the step of determining (20b, 20c) of a tuning underlying the music signal using the absolute tone height values of notes for segments of the music signal.

18. Method in accordance with claim 17, wherein the step of determining the tuning comprises the following steps:

forming (20b) a multitude of frequency differences from the tone height values of the music signal to obtain a frequency difference coordinate system;

determining (20c) the absolute tuning underlying the music signal, using the frequency difference coordinate system and using a plurality of stored tuning coordinate systems (40) by means of a compensational calculation.

19. Method in accordance with claim 18, wherein the step of determining (20) of the tone height comprises a step of quantizing (20d) of the absolute tone height value on the basis of the absolute tuning and a reference standard tone, to obtain one note per segment.

20. Method in accordance with one of the preceding claims, wherein the step of segmenting (16) comprises the following step:

transforming (16a) of the time length of tones into standardized tone lengths by histogramming (16b) the time length and identifying (16c) a fundamental note length such that the time lengths of the tones may be indicated as integer multiples or integer fractions of the fundamental note length, and quantizing (16c) of the time lengths of the tones to the next integer multiple or the next integer fraction to obtain a quantized note length.

21. Method in accordance with claim 20, wherein the step of segmenting (16) further includes a step of determining (16e) a bar from the quantized note lengths by examining whether succeeding notes may be grouped to a bar scheme.

22. Method in accordance with claim 21, further comprising the following step:

examining (60) a sequence of notes representing the music signal, with each note being specified by a start, a length, and a tone height with respect to compositional rules and marking a note, which is not compatible with the compositional rules.

23. Method for referencing a music signal (70) in a database (78) comprising a note-based description of a plurality of database music signals, comprising the following steps:

transferring (72) the music signal into the note-based description (74) in accordance with one of the claims 1 to 22;

comparing (76) the note-based description (74) of the music signal with the note-based description of the plurality of database music signals in the database (78);

making (76) a statement with respect to the music signal (70) on the basis of the step of comparing.

24. Method in accordance with claim 23, wherein the note-based description for the database music signals has an MIDI-format, with a tone start and a tone end being specified as a function of time, and wherein, prior to the step of comparing, the following steps are carried out:

forming differential values between two adjacent notes of the music signal to obtain a difference note sequence;

forming differential values between two adjacent notes of the note-based description of the database music signal, and

wherein, in the step of comparing, the differential note sequence of the music signal is compared with the differential note sequence of a database music signal.

25. Method in accordance with claim 23 or 24, wherein the step of comparing (76) is carried out using a DNA sequencing algorithm and, in particular, using the Boyer-Moore algorithm.

26. Method in accordance with one of claims 23 to 25, wherein the step of making a statement comprises the identifying of the identity of the music signal and of a database music signal, if the note-based description of the database music signal and the note-based description of the music signal are identical.

27. Method in accordance with one of claims 23 to 25, wherein the step of making a statement with respect to the music signal identifies a similarity between the music signal (20) and a database music signal, unless all tone heights and/or tone lengths of the music signal match with tone heights and/or tone lengths the database music signal.

28. Method in accordance with one of claims 23 to 26, wherein the note-based description comprises a rhythm description and wherein, in the step of comparing (76), a comparison of the rhythms of the music signal and of the database music signal is carried out.

29. Method in accordance with one of claims 23 to 28, wherein the note-based description comprises a tone height description and wherein, in the step of comparing (76), the tone heights of the music signal are compared with the tone heights of a database music signal.

30. Method in accordance with one of claims 25 to 29, wherein, in the step of comparing (26), insert, replace or delete operations are carried out with the note-based description (74) of the music signal (70) and wherein, in the step of making a statement, a similarity between the music signal (70) and a database music signal on the basis of the number of insert, replace or delete operations is identified, which are required to achieve a greatest possible matching between the note-based description (74) of the music signal (70) and the note-based description of a database music signal.

31. Apparatus for transferring a music signal into a note-based description, comprising:

means for generating (10) a frequency-time representation of the music signal, with the frequency-time representation comprising coordinate tuples, with a coordinate tuple including a frequency value and a time value, wherein the time value indicating the time of occurrence of the assigned frequency in the music signal;

means for calculating (12) a fit function as a function of time, the course of which is determined by the coordinate tuples of the frequency-time representation;

means for determining (14) at least two adjacent extreme values of the fit function;

means for time-segmenting (16) the frequency-time representation on the basis of the determined extreme values, with one segment being limited by two adjacent extreme values of the fit function, with the time length of the segment indicating a time length of a note assigned to this segment; and

means for determining (20) a tone height of the note for the segment using coordinate tuples in the segment.

32. Apparatus for referencing a music signal (70) in a database (78), comprising a note-based description of a plurality of database music signals, comprising:

means for transferring (72) the music signal into a note-based description (74) by a method in accordance with one of the patent claims 1 to 22;

means for comparing (76) the note-based description (74) of the music signal with the note-based description of the plurality of database music signals in the data bank (78); and

means for making (76) a statement with respect to the music signal (70) and the basis of the step of comparing.

Revendications

1. Procédé pour convertir un signal musical en une description basée sur des notes, aux étapes suivantes consistant à :

générer (10) une représentation fréquence-temps du signal musical, la représentation fréquence-temps présentant des uplets de coordonnées, un uplet de coordonnées comportant une valeur de fréquence et une valeur de temps, la valeur de temps indiquant le moment de l'occurrence de la fréquence associée dans le signal de musique ;
calculer (12) une fonction d'adaptation en fonction du temps dont l'évolution est déterminée par l'uplet de coordonnées de la représentation fréquence-temps ;
déterminer (14) au moins deux extrêmes adjacents de la fonction d'adaptation ;
segmenter dans le temps (16) la représentation fréquence-temps sur base des extrêmes déterminés, un segment étant limité par deux extrêmes adjacents de la fonction d'adaptation, la longueur dans le temps du segment indiquant une longueur dans le temps d'une note associée à ce segment ; et
déterminer (20) une hauteur tonale de la note pour le segment à l'aide d'uplets de coordonnées dans le segment.

2. Procédé selon la revendication 1, dans lequel la fonction d'adaptation est une fonction analytique, le dispositif (14) pour déterminer des extrêmes adjacents effectuant une différentiation de la fonction analytique et une détermination de points zéro.

3. Procédé selon la revendication 1 ou 2, dans lequel les valeurs extrêmes qui sont déterminées par le dispositif (14) sont des minimums de la fonction d'adaptation.

4. Procédé selon l'une des revendications précédentes, dans lequel la fonction d'adaptation est une fonction d'adaptation de polynôme du degré n, n étant supérieur à 2.

5. Procédé selon l'une des revendications précédentes, dans lequel est déterminée, à l'étape de segmentation (16), la longueur dans le temps d'une note, à l'aide d'une valeur d'étalonnage à partir de la distance dans le temps entre deux valeurs extrêmes adjacentes, la valeur d'étalonnage étant le rapport entre une longueur dans le temps prédéterminée d'un son, et une distance entre deux valeurs extrêmes déterminée pour le son à l'aide de la fonction d'adaptation.

6. Procédé selon la revendication 4 ou 5, dans lequel le degré de la fonction d'adaptation est déterminé d'avance à l'aide de sons prédéterminés de différentes longueurs connues et pour des fonctions d'adaptation de différents degrés, le degré étant utilisé à l'étape de calcul (12) où il est obtenu une coïncidence spécifiée entre les longueurs de son déterminées par des valeurs extrêmes adjacentes et les longueurs de son connues.

7. Procédé selon l'une des revendications 3 à 6, dans lequel il n'est segmenté, à l'étape de segmentation dans le temps (16), qu'à un minimum de la fonction d'adaptation dont la valeur de fréquence est différente de la valeur de fréquence d'un maximum adjacent d'au moins une valeur de seuil minimum-maximum, afin d'éliminer des minimums imaginaires.

8. Procédé selon l'une des revendications précédentes, dans lequel sont réalisées, à l'étape de génération (10), les étapes suivantes consistant à :

détecter (10c) l'occurrence dans le temps de flancs de signal dans le signal dans le temps ; déterminer (10d) une distance dans le temps entre deux flancs de signal détectés choisis et calculer une valeur de fréquence à partir de la distance dans le temps déterminés et associer la valeur de fréquence à un moment d'occurrence de la valeur de fréquence dans le signal musical, pour obtenir un uplet de coordonnées à partir de la valeur de fréquence et du moment d'occurrence pour cette valeur de fréquence.

9. Procédé selon la revendication 8, dans lequel est effectuée, à l'étape de détection (10c), une transformation de Hough.

10. Procédé selon l'une des revendications précédentes, dans lequel, à l'étape de génération (10), la représentation fréquence-temps est filtrée (10e), de sorte qu'il reste une bande de contour de hauteur tonale, et dans lequel il n'est tenu compte, à l'étape de calcul (12) d'une fonction d'adaptation, que des uplets de coordonnées dans la bande de contour de hauteur tonale.

11. Procédé selon l'une des revendications précédentes, dans lequel le signal musical est monophonique ou polyphonique à partie monophonique dominante.

12. Procédé selon la revendication 11, dans lequel le signal musical est une suite de notes chantée ou interprétée à l'aide d'un instrument.

13. Procédé selon l'une des revendications précédentes, dans lequel est effectuée, à l'étape (10) de génération d'une représentation fréquence-temps, une conversion de taux de balayage à un taux de balayage prédéterminé (10b).

14. Procédé selon l'une des revendications précédentes, dans lequel est effectuée, à l'étape (10) de génération d'une représentation fréquence-temps, une normalisation d'intensité sonore (10b) par multiplication par un facteur de modulation, qui est fonction de l'intensité sonore moyenne d'un segment et d'une intensité sonore maximale prédéterminée.

15. Procédé selon l'une des revendications précédentes, dans lequel est effectué, à l'étape de génération (10), un post-traitement spécifique à l'instrument (10f) de la représentation fréquence-temps, pour obtenir une représentation fréquence-temps spécifique à l'instrument, et

dans lequel, à l'étape du calcul (12) de la fonction d'adaptation, la représentation fréquence-temps spécifique à l'instrument sert de base.

16. Procédé selon l'une des revendications précédentes, dans lequel est utilisée, à l'étape de détermination (20) de la hauteur tonale par segment, la valeur moyenne des uplets de coordonnées dans un segment ou la valeur médiane des uplets de coordonnées dans le segment, la valeur moyenne ou la valeur médiane dans un segment indiquant une valeur

de hauteur tonale absolue de la note pour le segment.

17. Procédé selon la revendication 16, dans lequel l'étape de détermination (20) de la hauteur tonale présente l'étape consistant à déterminer (20b, 20c) une ambiance à la base du signal musical à l'aide des valeurs de hauteur tonale absolues de notes pour des segments du signal musical.

18. Procédé selon la revendication 17, dans lequel l'étape de détermination de l'ambiance présente les caractéristiques suivantes :

former (20b) une pluralité de différences de fréquence à partir des valeurs de hauteur tonale du signal musical, pour obtenir un système de coordonnées de différence de fréquence ;
déterminer (20c) l'ambiance absolue qui est à la base du signal musical, à l'aide du système de coordonnées de différence de fréquence et à l'aide d'une pluralité de systèmes de coordonnées d'ambiance mémorisés (40) au moyen d'un calcul de compensation.

19. Procédé selon la revendication 18, dans lequel l'étape de détermination (20) de la hauteur tonale présente une étape consistant à quantifier (20d) les valeurs de hauteur tonale absolues sur base de l'ambiance absolue et du diapason de référence, pour obtenir une note par segment.

20. Procédé selon l'une des revendications précédentes, dans lequel l'étape de segmentation (16) présente l'étape suivante consistant à :

transformer (16a) la longueur dans le temps de sons en longueurs de note normalisées par établissement d'un histogramme (16b) de la longueur dans le temps et fixation (16c) d'une longueur de note de base, de sorte que les longueurs dans le temps des sons puissent être indiquées comme multiples à nombre entier ou comme fractions à nombre entier de la longueur de note de base, et quantifier (16c] les longueurs dans le temps des sons sur le multiple à nombre entier le plus rapproché ou la fraction à nombre entier la plus rapprochée, pour obtenir des longueurs de note quantifiées.

21. Procédé selon la revendication 20, dans lequel l'étape de segmentation (16) comporte, par ailleurs, une étape consistant à déterminer (16e) une cadence à partir des longueurs de note quantifiées par examen de si des notes successives peuvent être regroupées en un schéma de cadence.

22. Procédé selon la revendication 21, présentant, par ailleurs, l'étape suivante consistant à :

examiner (60) une suite des notes représentant le signal musical, chaque note étant spécifiée par le début, la longueur et la hauteur tonale, quant aux règles de composition, et repérer une note qui n'est pas compatible avec les règles de composition.

23. Procédé pour référencer un signal musical (70) dans une banque de données (78) présentant une description basée sur des notes d'une pluralité signaux musicaux de banque de données, aux étapes suivantes consistant à :

convertir (72) le signal musical en une description basée sur des notes (74) selon l'une des revendications 1 à 22 ;
comparer (76) la description basée sur des notes (74) du signal musical avec la description basée sur des notes de la pluralité de signaux musicaux de banque de données dans la banque de données (78) ; et
établir (76) un proposition concernant le signal musical (70) sur base de l'étape de comparaison.

24. Procédé selon la revendication 23, dans lequel la description basée sur des notes pour les signaux musicaux de banque de données a un format Midi, un début de son et une fin de son étant spécifiés en fonction du temps, et dans lequel sont effectuées, avant l'étape de comparaison, les étapes suivantes consistant à :

former des valeurs de différence entre deux notes adjacentes du signal musical, pour obtenir une suite de notes de différence ;
former des différences entre deux notes adjacentes de la description basée sur des notes du signal musical de banque de données , et
dans lequel, à l'étape de comparaison, la suite de notes de différence du signal musical est comparée à la suite de notes de différence d'un signal musical de banque de données.

25. Procédé selon la revendication 23 ou 24, dans lequel l'étape de comparaison (76) est effectuée à l'aide d'un algorithme de séquençage d'ADN et, en particulier, à l'aide de l'algorithme de Boyer-Moore.

26. Procédé selon l'une des revendications 23 à 25, dans lequel l'étape d'établissement d'une proposition présente la constatation de l'identité du signal musical (70) et d'un signal musical de banque de données si la description basée sur des notes du signal musical de banque de données et la description basée sur des notes du signal musical sont

identiques.

27. Procédé selon l'une des revendications 23 à 25, dans lequel l'étape d'établissement d'une proposition sur le signal musical constate une similitude entre le signal musical (70) et un signal musical de banque de données concernant le signal musical lorsque pas toutes les hauteurs tonales et/ou les longueurs de son du signal musical coïncident avec les hauteurs tonales et/ou les longueurs de son du signal musical de banque de données.

28. Procédé selon l'une des revendications 23 à 27, dans lequel la description basée sur des notes présente une description de rythme, et dans lequel est effectuée, à l'étape de comparaison (76), une comparaison des rythmes du signal musical et du signal musical de banque de données.

29. Procédé selon l'une des revendications 23 à 28, dans lequel la description basée sur des notes présente une description de hauteur tonale, et dans lequel, à l'étape de comparaison (76), les hauteurs tonales du signal musical sont comparées aux hauteurs tonales d'un signal musical de banque de données.

30. Procédé selon l'une des revendications 25 à 29, dans lequel sont effectuées, à l'étape de comparaison (26), des opérations d'insertion, de remplacement ou d'effacement à l'aide de la description basée sur des notes (74) du signal musical (70), et dans lequel est constatée, à l'étape d'établissement d'une proposition, une similitude entre le signal musical (70) et un signal musical de banque de données sur base du nombre d'opérations d'insertion, de remplacement ou d'effacement nécessaires pour obtenir une coïncidence la plus grande possible entre la description basée sur des notes (74) du signal musical (70) et la description basée sur des notes d'un signal musical de banque de données.

31. Dispositif pour convertir un signal musical en une description basée sur des notes, aux caractéristiques suivantes :

un dispositif destiné à générer (10) une représentation fréquence-temps du signal musical, la représentation fréquence-temps présentant des uplets de coordonnées, un uplet de coordonnées comportant une valeur de fréquence et une valeur de temps, la valeur de temps indiquant le moment de l'occurrence de la fréquence associée dans le signal musical ;
un dispositif destiné à calculer (12) une fonction d'adaptation en fonction du temps dont l'évolution est déterminée par les uplets de coordonnées de la représentation fréquence-temps ;
un dispositif destiné à déterminer (14) au moins deux extrêmes adjacents de la fonction d'adaptation ;
un dispositif destiné à segmenter dans le temps (16) la représentation fréquence-temps sur base des extrêmes déterminés, un segment étant limité par deux extrêmes adjacents de la fonction d'adaptation, la longueur dans le temps du segment indiquant une longueur dans le temps d'une note associée à ce segment ; et
un dispositif destiné à déterminer (20) une hauteur tonale de la note pour ce segment à l'aide d'uplets de coordonnées dans le segment.

32. Dispositif pour référencer un signal musical (70) dans une banque de données (78) présentant une description basée sur des notes d'une pluralité de signaux musicaux de banque de données, aux caractéristiques suivantes :

un dispositif destiné à convertir (72) le signal musical en une description basée sur des notes (74) par un procédé selon l'une des revendications 1 à 22 ;
un dispositif destiné à comparer (76) la description basée sur des notes (74) du signal musical à la description basée sur des notes de la pluralité de signaux musicaux de banque de données dans la banque de données (78) ; et
un dispositif destiné à établir (76) une proposition concernant le signal musical (70) sur base de l'étape de comparaison.

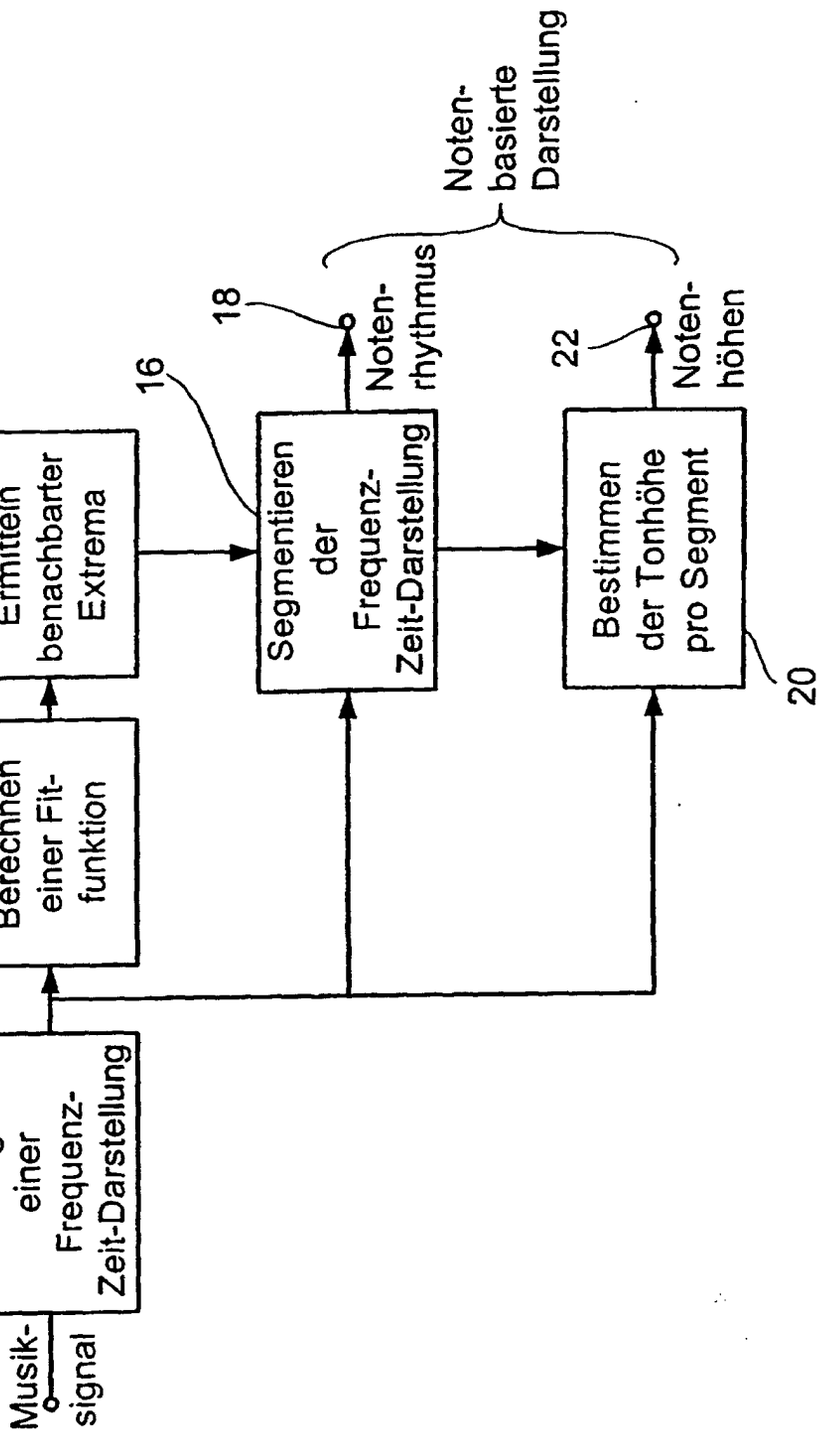
Musik-signal
10
Erzeugen einer Frequenz-Zeit-Darstellung
12
Berechnen einer Fit-funktion
14
Ermitteln benachbarter Extrema
16
Segmentieren der Frequenz-Zeit-Darstellung
18
Noten-rhythmus
20
Bestimmen der Tonhöhe pro Segment
22
Noten-höhen
Noten-basierte Darstellung

Fig. 1

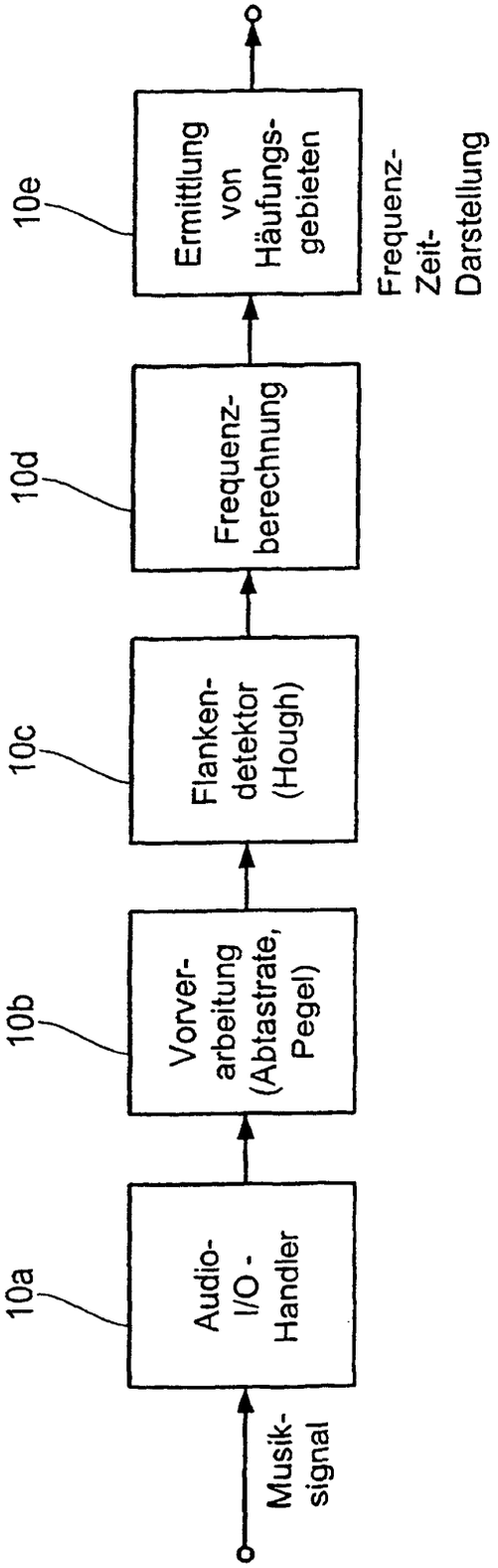
Musik-signal
10a
Audio-I/O-Handler
10b
Vorver-arbeitung (Abtastrate, Pegel)
10c
Flanken-detektor (Hough)
10d
Frequenz-berechnung
10e
Ermittlung von Häufungs-gebieten
Frequenz-Zeit-Darstellung

Fig. 2

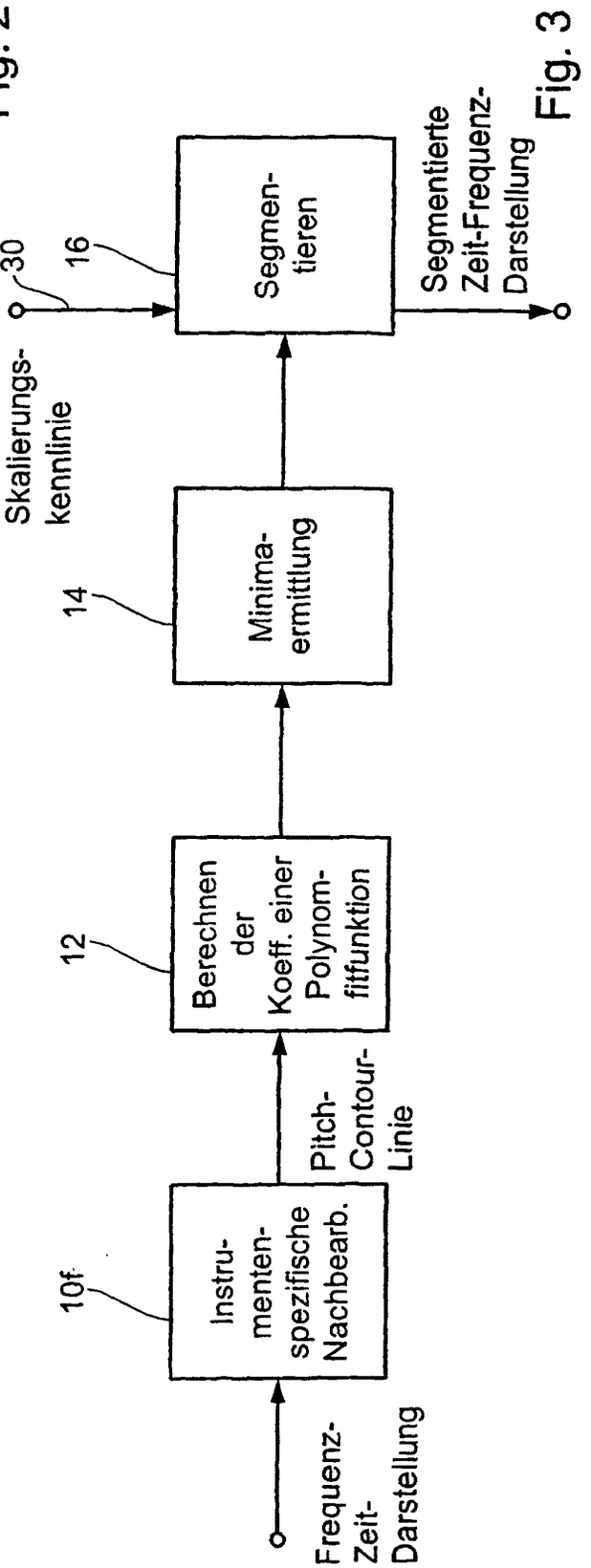
Frequenz-Zeit-Darstellung
10f
Instru-menten-spezifische Nachbearb.
Pitch-Contour-Linie
12
Berechnen der Koeff. einer Polynom-fitfunktion
14
Minima-ermittlung
Skalierungs-kennlinie
30
16
Segmen-tieren
Segmentierte Zeit-Frequenz-Darstellung

Fig. 3

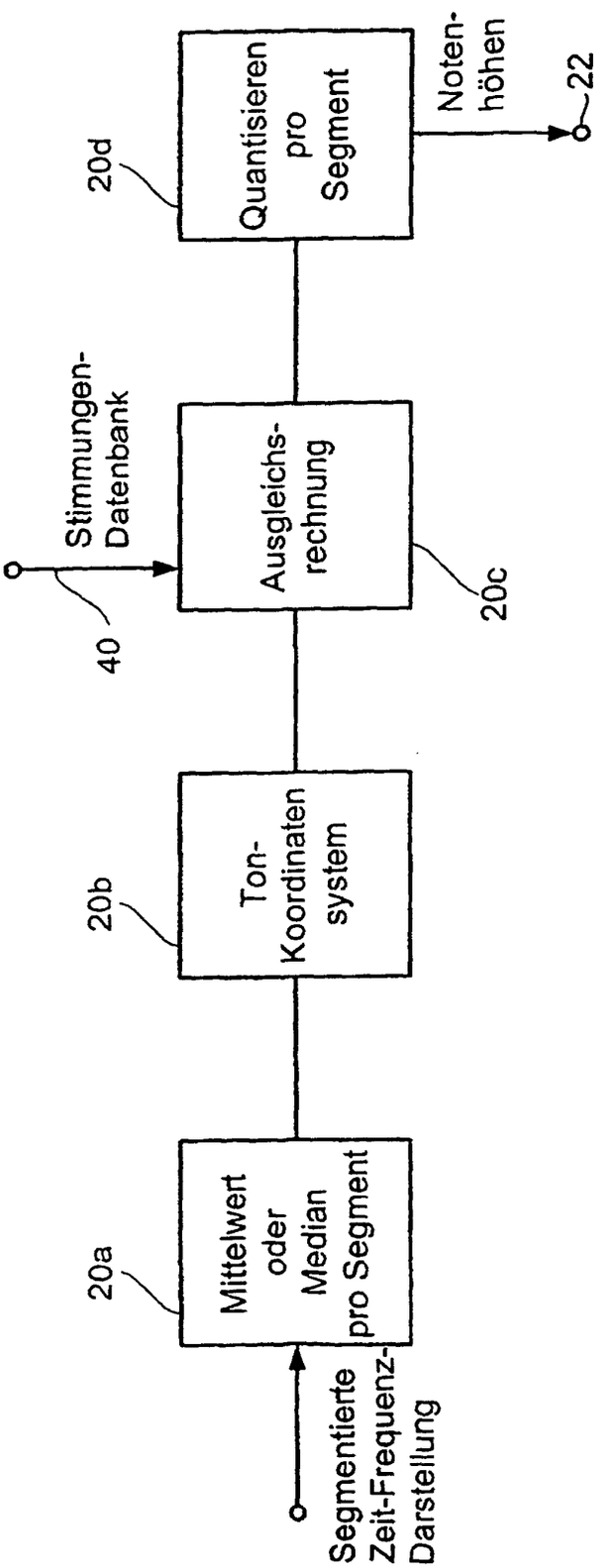
Segmentierte Zeit-Frequenz-Darstellung
20a
Mittelwert oder Median pro Segment
20b
Ton-Koordinaten system
40
Stimmungen-Datenbank
20c
Ausgleichs-rechnung
20d
Quantisieren pro Segment
Noten-höhen
22

Fig. 4

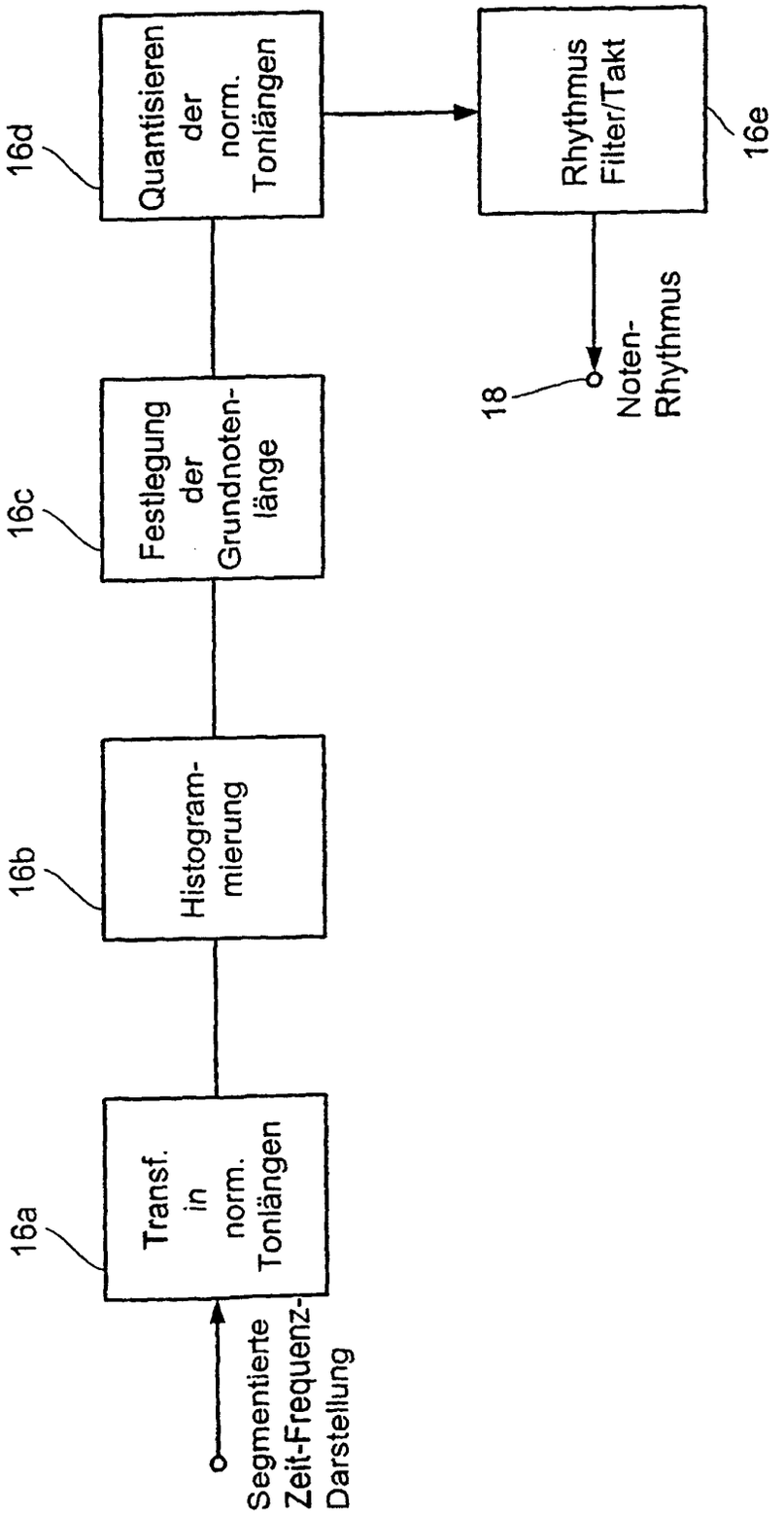
Segmentierte Zeit-Frequenz-Darstellung
16a
Transf. in norm. Tonlängen
16b
Histogram-mierung
16c
Festlegung der Grundnoten-länge
16d
Quantisieren der norm. Tonlängen
16e
Rhythmus Filter/Takt
18
Noten-Rhythmus

Fig. 5

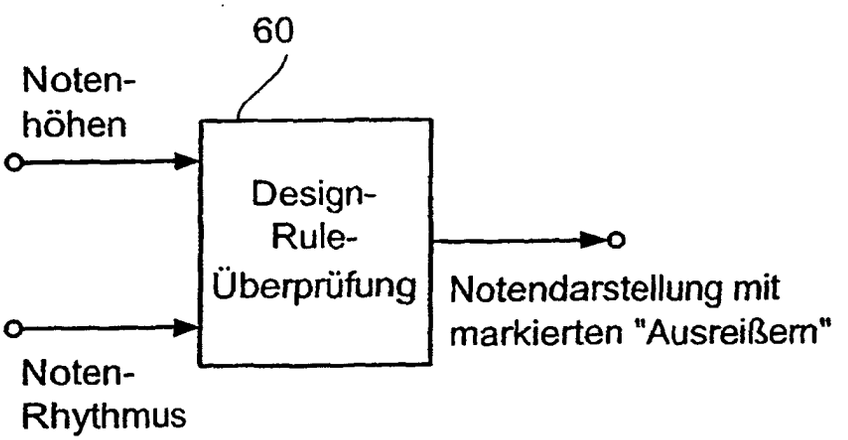
60
Noten-
höhen
Design-
Rule-
Überprüfung
Notendarstellung mit
markierten "Ausreißern"
Noten-
Rhythmus

Fig. 6

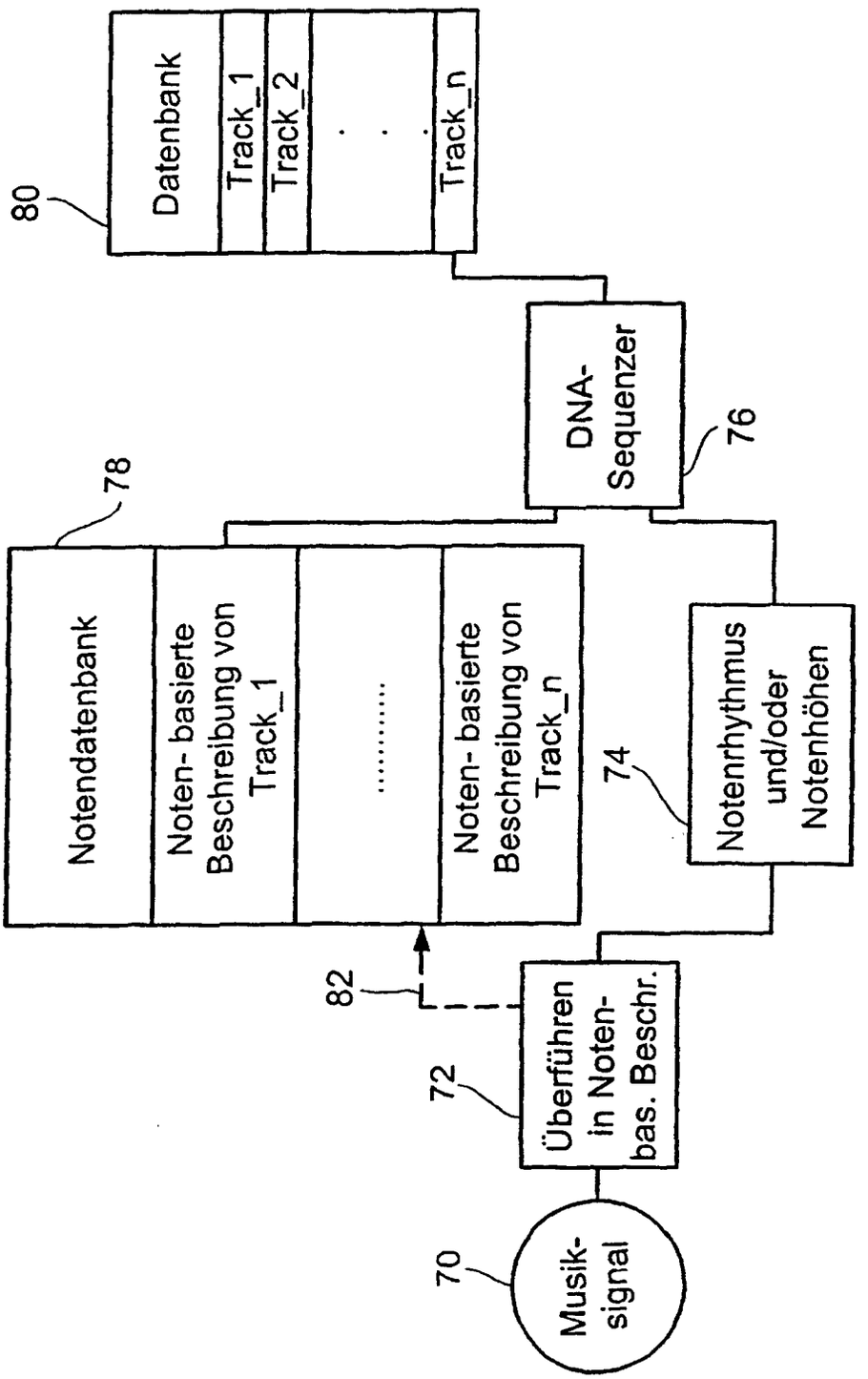
Musik-signal
70
Überführen in Noten-bas. Beschr.
72
82
Notendatenbank
Noten- basierte Beschreibung von Track_1
.....
Noten- basierte Beschreibung von Track_n
78
74
Notenrhythmus und/oder Notenhöhen
DNA-Sequenzer
76
Datenbank
Track_1
Track_2
. . .
Track_n
80

Fig. 7

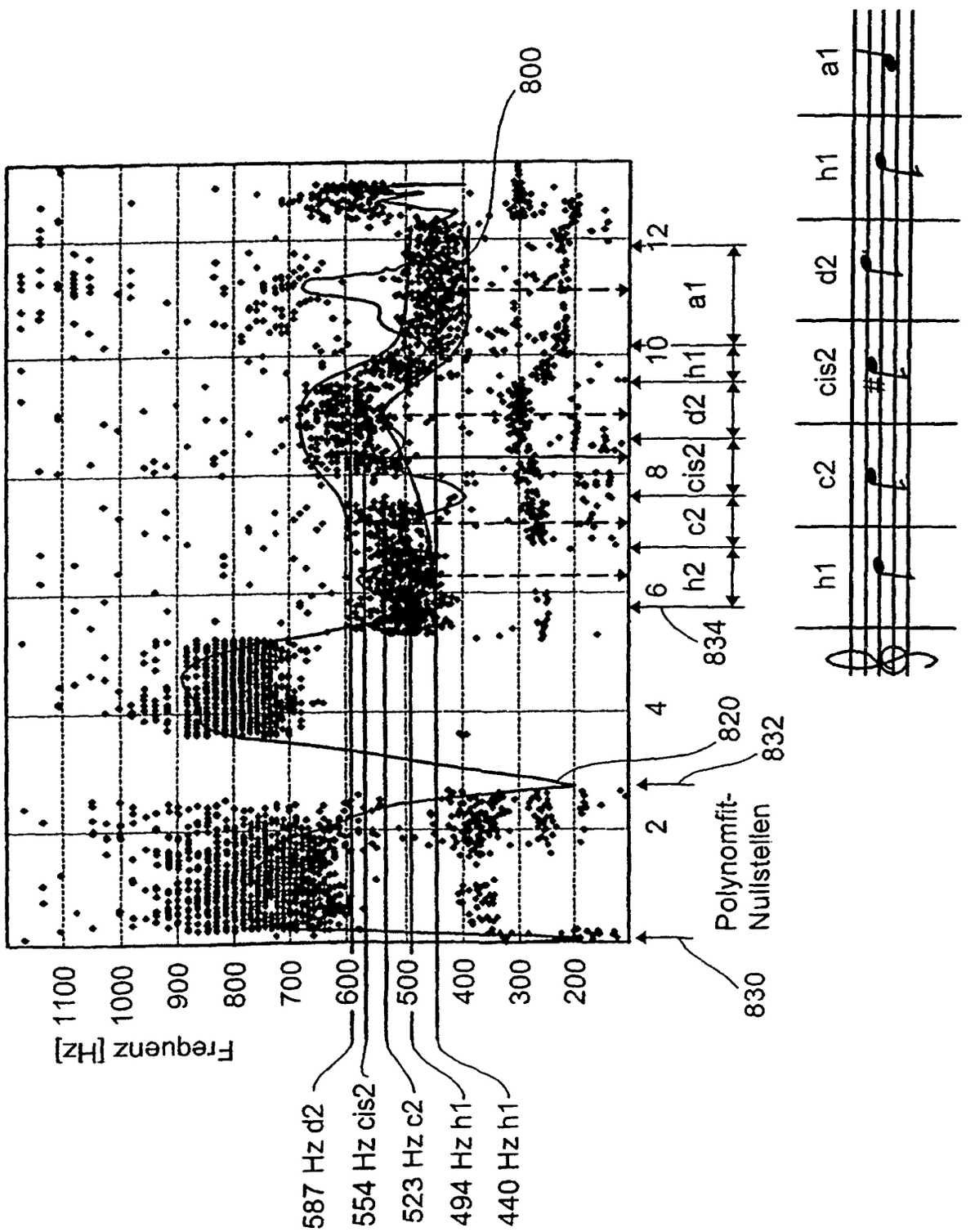
Frequenz [Hz]
1100
1000
900
800
700
600
500
400
300
200
587 Hz d2
554 Hz cis2
523 Hz c2
494 Hz h1
440 Hz h1
800
830
820
832
834
Polynomfit-Nullstellen
2
4
6
8
10
12
h2
c2
cis2
d2
h1
a1
h1
c2
cis2
d2
h1
a1

Fig. 8