

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
22. Dezember 2005 (22.12.2005)

PCT

(10) Internationale Veröffentlichungsnummer
WO 2005/122133 A1

(51) Internationale Patentklassifikation⁷: **G10H 1/00**

(21) Internationales Aktenzeichen: PCT/EP2005/002304

(22) Internationales Anmeldedatum:
4. März 2005 (04.03.2005)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
10 2004 028 720.1 14. Juni 2004 (14.06.2004) DE

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Hansastrasse 27c, 80686 München (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): **DERBOVEN, Claas** [DE/DE]; Thomas-Mann-Str. 4, 98693 Ilmenau (DE).
STREICH, Sebastian [DE/DE]; Langewiesener Str. 10 a, 98693 Ilmenau (DE). **CREMER, Markus** [DE/DE]; Herderstr. 4, 98693 Ilmenau (DE).

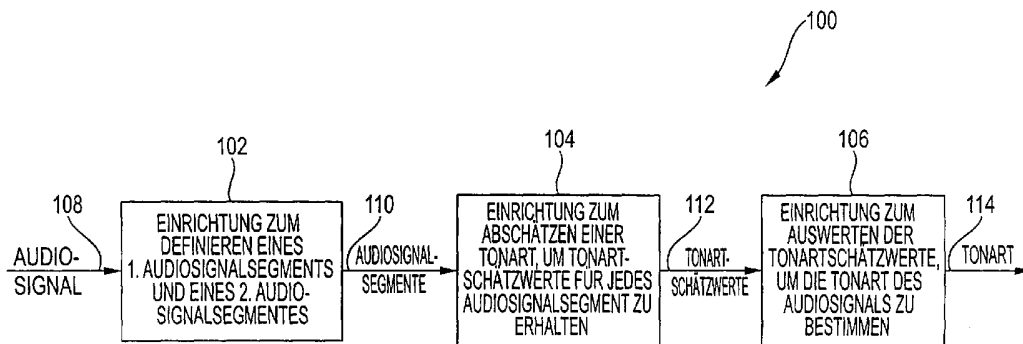
(74) Anwälte: **ZINKLER, Franz** usw.; SCHOPPE, ZIMMERMANN, STÖCKELER & ZINKLER, Postfach 246, 82043 Pullach bei München (DE).

(81) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ,

[Fortsetzung auf der nächsten Seite]

(54) Title: DEVICE AND METHOD FOR DETERMINING THE TONALITY OF AN AUDIO SIGNAL COMPRISING MUSICAL INFORMATION

(54) Bezeichnung: VORRICHTUNG UND VERFAHREN ZUM BESTIMMEN EINER TONART EINES AUDIOSIGNALS MIT MUSIKINFORMATIONEN



108 ... AUDIO SIGNAL

102 ... UNIT FOR DEFINING A 1st AUDIO SIGNAL SEGMENT AND A 2nd AUDIO SIGNAL SEGMENT

110 ... AUDIO SIGNAL SEGMENTS

104 ... UNIT FOR ESTIMATING A TONALITY, TO OBTAIN ESTIMATED TONALITY VALUES FOR EACH AUDIO SIGNAL SEGMENT

112 ... ESTIMATED TONALITY VALUES

106 ... UNIT FOR EVALUATING THE ESTIMATED TONALITY VALUES, TO DETERMINE THE TONALITY OF THE AUDIO SIGNAL

114 ... TONALITY

(57) Abstract: The invention relates to a device (100) for determining the tonality (114) of an audio signal (108) comprising musical information. Said device comprises a unit (102) for defining a first audio signal segment (ASS1) and a second audio signal segment (ASS2), the first segment (ASS1) and the second segment (ASS2) overlapping and having different durations. The device (100) also comprises a unit (104) for estimating a tonality for each audio signal segment (ASS1, ASS2), in order to obtain a first estimated tonality value for the first audio signal segment (ASS1) and a second estimated tonality value for the second audio signal segment (ASS2). In addition, the device (100) comprises a unit (106) for evaluating the estimated tonality values (112), in order to determine the tonality (114) of the audio signal (108), based on the first estimated tonality value and the second estimated tonality value. A device of this type provides an improved temporal resolution for the determination of the tonality of an audio signal.

[Fortsetzung auf der nächsten Seite]

WO 2005/122133 A1



TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(84) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL,

Veröffentlicht:

— mit internationalem Recherchenbericht

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(57) Zusammenfassung: Eine Vorrichtung (100) zum Bestimmen einer Tonart (114) eines Audiosignals (108) mit Musikinformationen umfasst eine Einrichtung (102) zum Definieren eines ersten Audiosignalsegments (ASS1) und eines zweiten Audiosignalsegments (ASS2), wobei sich das erste Audiosignalsegment (ASS1) und das zweite Audiosignalsegment (ASS2) überlappen und unterschiedliche Zeitdauern haben. Ferner umfasst die Vorrichtung (100) zum Bestimmen eine Einrichtung (104) zum Abschätzen einer Tonart für jedes Audiosignalsegment (ASS1, ASS2), um einen ersten Tonartschätzwert für das erste Audiosignalsegment (ASS1) und einen zweiten Tonartschätzwert für das zweite Audiosignalsegment (ASS2) zu erhalten. Außerdem hat die Vorrichtung (100) zum Bestimmen eine Einrichtung (106) zum Auswerten der Tonartschätzwerte (112), um auf der Basis des ersten Tonartschätzwertes und des zweiten Tonartschätzwertes die Tonart (114) des Audiosignals (108) zu bestimmen. Eine derartige Vorrichtung bietet den Vorteil eine verbesserte zeitliche Auflösung bei der Bestimmung der Tonart eines Audiosignals zu ermöglichen.

- 1 -

Vorrichtung und Verfahren zum Bestimmen einer Tonart eines
Audiosignals mit Musikinformatoren

Beschreibung

5

Die vorliegende Erfindung bezieht sich auf das technische
Gebiet der Tonarterkennung eines Musikssignals und insbesondere
bezieht sich die vorliegende Erfindung auf eine Vorrichtung und
ein Verfahren einer Tonartbestimmung eines Musikssignals, das
10 eine verbesserte zeitliche Auflösung des Musiksignals bietet.

In den letzten Jahren hat sich, bedingt durch die starke Auf-
splitterung des Musikgeschmacks, eine Vielzahl von unterschied-
lichen Musikrichtungen und -stilen ergeben. Zu diesen unter-
15 schiedlichen Musikstilen und -richtungen wurde in der letzten
Zeit eine sehr große und mittlerweile teils unüberschaubare
Menge von neuen Musikstücken durch die verschiedensten Inter-
preten erstellt, so dass nunmehr eine Person, die nicht 100
prozentig auf den aktuellsten musikalisch Stand ist, teilweise
20 erhebliche Schwierigkeiten hat, den Überblick über die ver-
schiedensten Stücke bzw. die verschiedensten Musikrichtungen
und -Stile zu behalten. Um dieses Problem etwas zu entschärfen,
werden Systeme und Verfahren entwickelt, um eine gewisse Klas-
sifizierung der Musikstücke anhand von vorgegebenen Kriterien
25 durchführen zu können, um einer solchen nicht 100 prozentig ak-
tuellsten Stand befindlichen Person, eine "vorgefilterte" Menge
von Musikstücken zur Auswahl zu stellen, die beispielsweise ein-
nem bestimmten musikalischen Geschmack der Person nahe kommt.
Für eine Klassifikation von derartigen Musikstücken, die dem
30 musikalischen Geschmack der Person entsprechen, kann beispiels-
weise das Kriterium der Tonart eines Musikstücks verwendet wer-
den, was insbesondere für die klassische Musik eine gute Hilfe-
stellung für eine Auswahl der verschiedenen Musikrichtungen und
-Stile ist.

35

Problematisch bei der Bestimmung der Tonart eines Musikstücks,
das in der Form eines Audiosignals mit Musikinformatoren vor-

- 2 -

liegt, ist in diesem Zusammenhang allerdings zu nennen, dass die wenigsten Musikstücke durchgehend nur eine einzige Tonart aufweisen. Vielmehr treten in einzelnen Segmenten des Musikstücks verschiedene Tonarten auf, so dass innerhalb des Musikstücks, das heißt, innerhalb des Audiosignals mit den Musikin-
5 formationen, Abschnitte auftreten, denen eine unterschiedliche Tonart zugrunde liegt. Dies wiederum stellt bei der Ermittlung der Tonart des gesamten Musikstücks ein Problem dar. Dieses Problem besteht insbesondere darin, die Tonart eines Musik-
10 stücks meist durch eine Transformation des üblicherweise in Zeitbereichsdarstellung vorliegenden Audiosignals in den Frequenzbereich erfolgt, wobei eine sehr "scharfe" spektrale Auflösung nur dann erreicht wird, wenn ein möglichst langes zeitliches Audiosignal verwendet wird. Tritt innerhalb des trans-
15 formierten Audiosignals lediglich eine Tonart auf, wäre eine Tonartbestimmung durch ein solches Verfahren nicht problematisch. Da, wie zuvor ausgeführt, in den meisten Musikstücken jedoch Tonartwechsel auftreten, kann bei einer Transformation eines Audiosignals oder eines Segments des Audiosignals, das
20 einen Tonartwechsel umfasst, meist keine eindeutige Entscheidung getroffen werden, da durch mehrere, in dem Audiosignalsegment auftretende Tonarten, eine eindeutige Zuordnung des untersuchten Audiosignalsegments zu einer einzigen Tonart erschwert oder gar unmöglich wird. In diesem Fall treten dann meist Fehler bei der Tonarterkennung auf, die dann meist zu einer deutlichen Behinderung der Einsetzbarkeit eines derartigen Ansatzes beiträgt.

Um diesem Problem Rechnung zu tragen, kann versucht werden, die
30 vom Zeitbereich in den Frequenzbereich zu transformierenden Audiosignalsegmente aus dem Audiosignal sehr kurz zu gestalten, so dass möglichst ein Wechsel der Tonart innerhalb eines Audio-segments vermieden wird. Ein derartiger Ansatz weist allerdings den Nachteil auf, dass durch zeitlich sehr kurze Audiosignalsegmente nur eine "unscharfe" spektrale Auflösung der in dem
35 Audiosignalsegment enthaltenden Frequenzanteile möglich ist,

- 3 -

wodurch wiederum die Zuordnung einer Tonart zu dem Audiosignal erschwert wird.

Eine Ermittlung der Tonart des Audiosignalsegments kann in diesem Zusammenhang beispielsweise durch das von David Temperley in dessen Buch „The cognition of basical musical structues“, The MIT Press, 2001, Seiten 173 bis 187 vorgeschlagene Verfahren erfolgen, in welchem eine zeitliche Auftrittsdauer von Tönen in einer Oktave bestimmt und mit verschiedenen Tonart-Referenzvektoren korreliert werden, um aufgrund eines maximalen Korrelationskoeffizienten die dem Tonart-Referenzvektor entsprechende Tonart als Tonart des Audiosignalsegments zu bestimmen.

Die Problematik bei der Bestimmung einer möglichst "scharfen" Frequenzauflösung und zugleich einem Vermeiden von Tonartwechseln innerhalb der zeitlichen Dauer des zu untersuchenden Audiosignalsegments besteht somit darin, eine Zeitdauer zu finden, die für alle zu untersuchenden Musikstücken gute Ergebnisse für die Tonartbestimmung liefert. Ist die Zeitdauer des zu untersuchenden Signalsegments allerdings zu kurz, sind die Resultate, wie oben dargestellt, erstens aufgrund der unscharfen Frequenzauflösung unzuverlässig und zweitens kann durch eine kurze zeitliche Dauer keine präzise statistische Auswertung über die Auftrittsdauer von einzelnen Tönen einer Oktave innerhalb des auszuwertenden Audiosignalsegments erfolgen, was wiederum die Tonartbestimmung auf der Basis einer kurzen zeitlichen Dauer des Audiosignalsegments behindert. Sind jedoch andererseits die zeitlichen Dauern der Audiosignalsegmente zu lange, kann ein Tonartwechsel, der innerhalb des Audiosignalsegments erfolgt, nicht detektiert und damit für die Bestimmung einer Tonart des gesamten Musikstücks nicht berücksichtigt werden.

Der vorliegenden Erfindung liegt daher die Aufgabe zugrunde, eine Vorrichtung und ein Verfahren zum Bestimmen einer Tonart eines Audiosignals mit Musikinformationen zu schaffen, das in

- 4 -

Bezug auf ein zeitliches Auflösungsverhalten verbesserte Charakteristika aufweist.

5 Diese Aufgabe wird durch eine Vorrichtung gemäß Anspruch 1 sowie ein Verfahren gemäß Anspruch 13 gelöst.

Die vorliegende Erfindung schafft eine Vorrichtung zum Bestimmen einer Tonart eines Audiosignals mit Musikinformationen, die folgende Merkmale aufweist:

10

eine Einrichtung zum Definieren eines ersten Audiosignalsegments und eines zweiten Audiosignalsegments, wobei sich das erste Audiosignalsegment und das zweite Audiosignalsegment überlappen und unterschiedliche Zeitdauern haben;

15

eine Einrichtung zum Abschätzen einer Tonart für jedes Audiosignalsegment, um einen ersten Tonartschätzwert für das erste Audiosignalsegment und einen zweiten Tonartschätzwert für das zweite Audiosignalsegment zu erhalten; und

20

eine Einrichtung zum Auswerten der Tonartschätzwerte, um auf der Basis des ersten Tonartschätzwerts und des zweiten Tonartschätzwerts die Tonart des Audiosignals zu bestimmen.

25 Ferner schafft die vorliegende Erfindung ein Verfahren zum Bestimmen einer Tonart eines Audiosignals mit Musikinformationen, das folgende Schritte aufweist:

30 Definieren eines ersten Audiosignalsegments und eines zweiten Audiosignalsegments, wobei sich das erste Audiosignalsegment und das zweite Audiosignalsegment überlappen und unterschiedliche Zeitdauern haben;

35 Abschätzen einer Tonart für jedes Audiosignalsegment, um einen ersten Tonartschätzwert für das erste Audiosignalsegment und einen zweiten Tonartschätzwert für das zweite Audiosignalsegment zu erhalten; und

- 5 -

Auswerten der Tonartschätzwerte, um auf der Basis des ersten Tonartschätzwerts und des zweiten Tonartschätzwerts die Tonart des Audiosignals zu bestimmen.

5

Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, dass durch das Definieren von mehreren Audiosignalsegmenten, die sich überlappen und unterschiedliche Zeitdauern haben, ein verbessertes zeitliches Auflösungsverhalten und damit einhergehend eine verbesserte Charakteristik einer Bestimmung einer Tonart eines Audiosignals möglich ist. Insbesondere dadurch, dass sich die Audiosignalsegmente überlappen, ist ein Signalanteil des Audiosignals somit in zwei verschiedenen Audiosignalsegmenten enthalten und wird somit bei der Bestimmung des Tonartschätzwerts für beide Audiosignalsegmente berücksichtigt. Außerdem wird durch die unterschiedlichen Zeitdauern der einzelnen Audiosignalsegmente ein unterschiedliches Auflösungsverhalten sowie eine unterschiedliche statistische Auswertung möglich, was zu einer Verbesserung der spektralen "Schärfe" und somit zu einer Verbesserung der Tonarterkennung für länger andauernde Audiosignalsegmente mit gleichbleibender Tonart führt. Dem gegenüber können durch das Abschätzen einer Tonart für das kürzere Audiosignalsegment gegenüber dem längeren Audiosignalsegment möglicherweise auftretende Tonartwechsel in dem längeren Audiosignalsegment erkannt werden. Werden nun für die einzelnen Audiosignalsegmente Tonartschätzwerte ermittelt, kann durch das Auswerten der einzelnen Tonartschätzwerte einerseits ein Rückschluss auf die Tonart des gesamten Audiosignals mit den Musikinformationen gezogen werden und andererseits auch das zeitliche Auftreten eines Tonartwechsels innerhalb des Audiosignals mit den Musikinformationen erkannt werden. Dieses Erkennen des Tonartwechsels kann insbesondere durch das Auswerten des Tonartschätzwerts für das kürzere Audiosignalsegment im Vergleich mit dem Tonartschätzwert für das längere Audiosignalsegment erfolgen.

- 6 -

Die vorliegende Erfindung bietet den Vorteil, eine verbesserte zeitliche Auflösung bei der Bestimmung der Tonart eines Audiosignals zu ermöglichen und dadurch eine verbesserte Charakteristik einer Vorrichtung zum Bestimmen einer Tonart eines Audiosignals zu schaffen. Insbesondere durch das Definieren des ersten Audiosignalsegments und des zweiten Audiosignalsegments, wobei sich das erste und zweite Audiosignalsegment überlappen und unterschiedliche Zeitdauern haben, lässt sich nun eine verbesserte Auswertung der Tonart eines Musiksignals durchführen, derart, dass gegenüber dem Stand der Technik Tonartwechsel innerhalb des Audiosignals durch die Auswertung des kürzeren Audiosignalsegments im Vergleich zum längeren Audiosignalsegment erkennbar werden. Ferner lässt sich auch durch die Auswertung der Tonartschätzwerte für jedes Audiosignalsegment ein Rückschluss für die Tonart des Audiosignals mit den Musikinformationen ziehen. Hierzu kann beispielsweise eine Häufigkeitsverteilung oder Statistik der auftretenden Ausprägungen der Tonartschätzwerte (beispielsweise in Form von Zahlenwerten, die Tonarten repräsentieren) ermittelt werden und diejenige Tonart, die der am häufigsten auftretenden Ausprägungen eines Tonartschätzwertes entspricht, als Tonart des Audiosignals bestimmt werden.

Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend anhand der beiliegenden Zeichnungen näher erläutert. Es zeigen:

Fig. 1 ein Blockschaltbild eines Ausführungsbeispiels einer erfindungsgemäßen Vorrichtung zum Bestimmen einer Tonart eines Audiosignals;

Fig. 2A eine Matrixdarstellung, wie sie durch die Einrichtung zum Abschätzen einer Tonart gemäß einem weiteren Ausführungsbeispiel erstellt wird;

Fig. 2B eine vergrößerte Darstellung des Zeilenanfangs der in Fig. 2A dargestellten ersten Matrixzeile; und

Fig. 3 eine Kreisdiagramm-Darstellung der unter Verwendung von Ausführungsbeispielen der vorliegenden Erfindung erhaltenen Ergebnisse.

5

In der nachfolgenden Beschreibung werden für gleiche oder gleich wirkende Elemente gleiche oder ähnlich Bezugszeichen verwendet, wobei auf eine wiederholte Beschreibung dieser Elemente verzichtet wird.

10

Fig. 1 zeigt ein Blockschaltbild eines Ausführungsbeispiels der erfindungsgemäßen Vorrichtung 100 zum Bestimmen einer Tonart eines Audiosignals. Die Vorrichtung 100 zum Bestimmen einer Tonart eines Audiosignals weist eine Einrichtung 102 zum Definieren eines ersten Audiosignalsegments und eines zweiten Audiosignalsegments, eine Einrichtung 104 zum Abschätzen einer Tonart, um Tonartschätzwerte für jedes Audiosignalsegment zu erhalten und eine Einrichtung 106 zum Auswerten der Tonartschätzwerte, um die Tonart des Audiosignals zu bestimmen. Die Einrichtung 102 zum Definieren eines ersten Audiosignalsegments und eines zweiten Audiosignalsegments umfasst einen Eingang zum Empfangen eines Audiosignals 108 und einen Ausgang zum Ausgeben der definierten Audiosignalsegmente 110. Ferner umfasst die Einrichtung 104 zum Abschätzen einen Eingang zum Empfangen der von der Einrichtung 102 zum Definieren ausgegebenen Audiosignalsegmente 110 sowie einen Ausgang zum Ausgeben der erhaltenen Tonartschätzwerte 112. Die Einrichtung 106 zum Auswerten umfasst einen Eingang zum Empfangen der von der Einrichtung 104 zum Abschätzen ausgegebenen Tonartschätzwerte 112 und einen Ausgang zum Ausgeben der bestimmten Tonart 114 des Audiosignals 108.

Die Funktionsweise der in Fig. 1 als Blockschaltbild dargestellten Vorrichtung 100 zum Bestimmen einer Tonart lässt sich wie folgt beschreiben: Zuerst wird ein Audiosignal 108 der Einrichtung 102 zum Definieren eines ersten Audiosignalsegments und eines zweiten Audiosignalsegments zugeführt. Die Einrich-

35

- 8 -

tung 102 zum Definieren bildet dann aus dem Audiosignal ein erstes Audiosignalsegment und ein zweites Audiosignalsegment, wobei sich das erste Audiosignalsegment und das zweite Audiosignalsegment überlappen und unterschiedliche Zeitdauern haben.

5 Hierbei wird durch die Einrichtung 102 zum Definieren ein variabler Zeitrahmen gewählt und beispielsweise zum Definieren des ersten Audiosignalsegments unter Verwendung des Zeitrahmens die Informationen in dem Audiosignal praktisch "herausgestanzt", die innerhalb des Zeitrahmens liegen. Der Zeitrahmen ist dabei

10 durch einen Startzeitpunkt und einen dem Startzeitpunkt zeitlich nachgelagerten Endzeitpunkt definiert. Anschließend wird durch die Einrichtung 102 der Zeitrahmen derart verändert, dass der Startzeitpunkt gleich bleibt, jedoch der Endzeitpunkt auf einen zeitlich dem ersten Endzeitpunkt nachgelagerten zweiten

15 Endzeitpunkt gesetzt wird, so dass der neu erstellte (zweite) Zeitrahmen eine größere zeitliche Länge aufweist, als der zum Erstellen des ersten Audiosignalsegments verwendete Zeitrahmen. Unter Verwendung des neuen, zweiten Zeitrahmens wird nun aus dem Audiosignal wieder praktisch ein Segment "herausge-

20 stanzt", wobei nun die Informationen, die im ersten Audiosignalsegment enthalten sind, auch im zweiten Audiosignalsegment enthalten sind. Zusätzlich weist jedoch das zweite Audiosignalsegment Informationen auf, die im Audiosignal in der Zeitdauer zwischen dem ersten Endzeitpunkt und dem zweiten Endzeitpunkt

25 angeordnet sind. Auf diese Weise werden zwei Audiosignalsegmente erstellt, die sich zeitlich überlappen, jedoch unterschiedliche Zeitdauern haben. Weiterhin kann die Einrichtung 102 ausgebildet sein, um weitere Audiosignalsegmente zu definieren, beispielsweise derart, dass für ein drittes Audiosig-

30 nalsegment der Startzeitpunkt des (dritten) Zeitrahmens mit dem ersten und zweiten Startzeitpunkt identisch ist und der dritte Endzeitpunkt auf einen Zeitpunkt gelegt wird, der zeitlich nach dem zweiten Endzeitpunkt angeordnet ist. Hierdurch würde dann der Zeitrahmen für das dritte Audiosignalsegment wieder größer

35 als der zweite Zeitrahmen, wobei durch die identischen Startzeitpunkte wieder die Information in dem zweiten Audiosignalsegment auch vollständig in dem dritten Audiosignalsegment ent-

- 9 -

halten ist, das dritte Audiosignalsegment jedoch zusätzlich die Informationen des Audiosignals umfasst, die zwischen dem zweiten Endzeitpunkt und dem dritten Endzeitpunkt liegt. Alternativ kann die Einrichtung 102 zum Definieren auch ausgebildet sein, um den Startzeitpunkt auf einen zeitlich nachgelagerten (zweiten) Startzeitpunkt zu legen. In diesem Fall lässt sich somit praktisch ein "Weiterschieben" des Zeitrahmens ausführen, indem dann beispielsweise wiederum nach der vorstehend beschriebenen Methode die Zeitrahmen für die Bildung von Audiosignalsegmenten entsprechend erweitert bzw. verlängert wird.

Die auf diese Weise durch die Einrichtung 102 zum Definieren erhaltenen Audiosignalsegmente 110 werden dann der Einrichtung 104 zum Abschätzen einer Tonart zugeführt, wobei die Einrichtung 104 zum Abschätzen für jedes Audiosignalsegment je einen Tonartschätzwert bestimmt. Diese Bestimmung des Tonartschätzwertes kann beispielsweise auf die eingangs dargelegte Weise nach David Temperley erfolgen. Diese Tonartschätzwerte können entweder Nominalwerte sein, die bereits eine Tonart charakterisieren (beispielsweise B-Dur, D#-moll, F#-moll etc.) sein. Alternativ kann auch jeder Tonart ein Zahlenwert zugeordnet werden, wobei dann der Tonartschätzwert diesem Zahlenwert entspricht. In einer besonderen Ausführungsform kann auch eine Matrixstruktur der Tonartschätzwerte unter Berücksichtigung der Start- und Endzeitpunkte erfolgen, wie er unter Bezugnahme auf die Fig. 2A und 2B nachfolgend näher beschrieben ist.

Die Einrichtung 104 zum Abschätzen kann dann die Tonartschätzwerte 112 entweder direkt oder alternativ in Form der Matrix an die Einrichtung 106 zum Auswerten der Tonartschätzwerte weitergegeben werden. Die Einrichtung 106 zum Auswerten der Tonartschätzwerte kann auf dann ausgebildet sein, um die Tonart 114 des Audiosignals 108 zu bestimmen. Dieses Bestimmen der Tonart 114 des Audiosignals 108 kann beispielsweise durch eine Auswertung der Häufigkeit des Auftretens von Ausprägungen der Tonartschätzwerte in Form der vorstehend genannten Zahlenwerte für jede Tonart erfolgen, wobei dann diejenige Tonart als Tonart

- 10 -

114 des Audiosignals 108 bestimmt wird, die am häufigsten auftritt, bzw. deren entsprechender Zahlenwert am häufigsten in den Audiosignalsegmenten auftritt. Alternativ kann aber auch eine andere Auswertung der einzelnen Tonartschätzwerte als die
5 Häufigkeitsverteilungsauswertung erfolgen.

Weiterhin kann die Einrichtung 106 zum Auswerten der Tonartschätzwerte auch ausgebildet sein, um Tonartwechsel aus den Tonartschätzwerten für die einzelnen Audiosignalsegmente zu erkennen. Diese Tonartwechsel können dann insbesondere gut erkannt werden, wenn die Tonartschätzwerte in der vorstehend bezeichneten Matrixform der Einrichtung 106 zum Auswerten übergeben werden, da hierdurch besonders einfach durch eine Differenzbildung (bei Tonartschätzwerten in Zahlendarstellung) möglich ist und somit auf einfache Art und Weise erkannt werden kann, wo eine Änderung der Tonartschätzwerte in den Elementen der Matrix auftritt. Diese Änderung der Tonartschätzwerte kann dann beispielsweise dadurch bestimmt werden, dass ein Referenzelement der Matrix einen Referenztonartschätzwert hat und ein
10 zum Referenzelement in der Matrix benachbart angeordnetes Bezugselement einen Bezugselementtonartschätzwert hat, der sich vom Referenztonartschätzwert unterscheidet. Bei der Ausbildung der Tonartschätzwerte in Zahlenform lässt sich dann weiterhin eine Optimierung ermitteln, derart, dass ein Tonartwechsel nur
15 dann erkannt wird, wenn eine Differenz zwischen dem Referenztonartschätzwert und dem Bezugstonartschätzwert größer als ein vordefinierter Schwellwert ist. Dies ermöglicht eine präzisere Aussage über das tatsächliche Vorliegen eines Tonartwechsels, wobei dann möglicherweise Fehlklassifizierungen vermieden werden können, bei denen in einzelnen Elementen der Matrix (beispielsweise Zeilen oder Spalten der Matrix) kurzzeitig unterschiedlich Tonartschätzwerte auftreten. In diesem Zusammenhang kann alternativ auch als Kriterium verwendet werden, dass in
20 einer Zeile mehrere gleiche Tonartschätzwerte in aufeinanderfolgenden Matrix-Elementen der Zeile auftreten sollten, um eindeutig einen Tonartwechsel in dem Audiosignal zu ermitteln. Ein solches Kriterium entspricht dabei praktisch einer erforderli-

- 11 -

chen zeitlichen Mindestlänge, über die eine Tonart in dem Audiosignal auftreten sollte, um erkannt zu werden.

Fig. 2A zeigt eine Matrix mit einer Diagonalstruktur, wobei die obere Diagonalmatrix mit Werten gefüllt ist. Die Zeilen der Matrix entsprechen hierbei Audiosignalsegmenten mit einem identischen Startzeitpunkt t_{start} (beispielsweise in Sekunden) und die Spalten der Matrix 200 entsprechen Audiosignalsegmenten mit identischen Endzeitpunkten t_{end} (beispielsweise ebenfalls in Sekunden). Die Diagonalstruktur der Matrix 200 ergibt sich dann daraus, dass für einen Startzeitpunkt eine Folge von Audiosignalsegmenten ausgewählt wird, wobei der Endzeitpunkt der Audiosignalsegmente bei gleichem Startzeitpunkt immer weiter (zeitlich) nach hinten verschoben wird. Hierdurch wird eine Auswertung des Audiosignals unter Berücksichtigung der Start- und Endzeitpunkte t_{start} und t_{end} möglich, wie sie in Fig. 2B dargestellt ist. Bei einer Länge des ersten Audiosignalsegments ASS1 von 2s (d.h. zwischen den Zeitpunkten 0s und 2s) und einer Länge des zweiten Audiosignals ASS2 von 4 Sekunden (d.h. von 0s bis 4s) ist somit der Signalgehalt des Audiosignals im Bereich von 0s bis 2s auch im zweiten Audiosignalsegment ASS2 enthalten. Dies wird durch die Schraffur von links unten nach rechts oben des ersten Audiosignalsegments im Vergleich zur Schraffur von links oben nach rechts unten des zweiten Audiosignalsegments ASS2 in Fig. 2B deutlich. Weiterhin könnte beispielsweise durch die Einrichtung zum Definieren ein zusätzliches Signalsegment ASS2a definiert werden, das Signalanteile des Audiosignals vom Startzeitpunkt $t_{\text{start}}=0\text{s}$ bis zum Endzeitpunkt $t_{\text{end}}=6\text{s}$ umfasst, so wie es in Fig. 2B durch die gepunktete Schraffur dargestellt ist. Hierbei umfasst wiederum das zusätzliche Audiosignalsegment ASS2a die Information der ersten und zweiten Audiosignalsegmente ASS1 und ASS2 zwischen den Zeitpunkten 0s und 4s.

In einer weiteren Ausführungsform kann die Einrichtung zum Definieren ein drittes Audiosignalsegment ASS3 definieren, das Informationen des Audiosignals von Startzeitpunkt $t_{\text{start}}=2\text{s}$ bis

- 12 -

zum Endzeitpunkt $t_{\text{end}}=4\text{s}$ umfasst. Hieran anschließend kann durch die Einrichtung zum Definieren ein viertes Audiosignalsegment ASS4 definiert werden, das Informationen des Audiosignals im Zeitbereich zwischen dem Startzeitpunkt $t_{\text{start}}=2\text{s}$ und dem Endzeitpunkt $t_{\text{end}}=6\text{s}$ umfasst und somit die Information des Audiosignals aus dem dritten Audiosignalsegment ASS3 beinhaltet.

Durch einen derartigen Aufbau und den zeitlichen Versatz der einzelnen Audiosignalsegmente kann somit die in Fig. 2A dargestellte Matrix erstellt werden, bei der die obere Diagonalmatrix gefüllt wird. Dieses Füllen kann beispielsweise dadurch erfolgen, dass für jedes Audiosignalsegment ein Tonartschätzwert ermittelt und dieser Wert in das Element eingetragen wird, der einen nicht von vorausgehenden Audiosignalsegmenten überlappenden Zeitbereich entspricht. Unter Bezugnahme auf die Fig. 2B kann dann beispielsweise der Tonartschätzwert des ersten Audiosignalsegments in das Element der Matrix eingetragen werden, das die Zeitspanne von 0 bis 2 Sekunden umfasst. Hierbei ist es unerheblich, ob der Tonartschätzwert flächig (d.h. als Fläche bzw. Flächenschraffur) oder als einzelner Zahlenwert in das Matrix-Element eingetragen wird; es ist lediglich die Zuordnung eines Tonartschätzwerts zu dem Intervall zwischen 0 Sekunden als Startzeitpunkt und 2 Sekunden als Endzeitpunkt des ersten Audiosignalsegments ASS1 relevant. Ferner kann der Audiosignalschätzwert für das zweite Audiosignalsegment ASS2 in dasjenige Element der Matrix eingetragen werden, das in der ersten Zeile zwischen den Zeitpunkten 2 Sekunden und 4 Sekunden liegt, obwohl das zweite Audiosignalsegment ASS2 Informationen umfasst, die im Bereich von 0 bis 2 Sekunden liegen. Hierdurch kann sichergestellt werden, dass die Information von vorhergehenden Audiosignalsegmenten (in diesem Fall dem ersten Audiosignalsegment ASS1) nicht verloren geht und zugleich die Information des zweiten Audiosignalsegments ASS2 angemessen in der Matrixdarstellung aufgenommen werden kann. Analoges gilt auch für das zusätzliche Audiosignalsegment ASS2a, das dritte Audiosignalsegment ASS3 und das vierte Audiosignalsegment ASS4 und so weiter.

- 13 -

Eine solche Eintragung der unterschiedlichen Tonartschätzwerte für die einzelnen Audiosignalsegmente kann dann in einer vollständigen Darstellung der in Fig. 2A dargestellten oberen Dreiecksmatrix entnommen werden. Hierbei ist ersichtlich, dass im Bereich von etwa 20 bis 40 Sekunden ein Tonartwechsel stattfindet, was durch den im Vergleich zu den umliegenden Tonartschätzwerten helleren Farbwert gekennzeichnet ist. Hierbei wird auch ersichtlich, dass aufgrund des durchgehenden hellen Bereichs von der Diagonale der Matrix bis zum oberen Rand ein Tonartwechsel erkennbar ist, und insbesondere dieser Tonartwechsel durch scharfe Kanten in der Matrix, wie beispielsweise der Kante 202, eindeutig ist. Diese scharfen Kanten zeichnen sich dann insbesondere dadurch aus, dass in übereinanderliegenden Zeilen zu gleichen Endzeitpunkten t_{end} ein Wechsel des Tonartschätzwertes zu den nachfolgenden Tonartschätzwerten in jeder Zeile auftritt.

Weiterhin lässt sich beispielsweise aus dem rechten oberen Element der Matrix 204 ein Erstschätzwert für die Tonart des Audiosignals bestimmen, indem die Anteile des kompletten Audiosignals berücksichtigt sind. Das Bereitstellen eines solchen Erstschätzwerts für die Tonart des Audiosignals bietet den Vorteil, bereits eine schnelle Grobklassifikation über die zu erwartende Tonart des Musikstücks bereitstellen zu können, ohne numerische aufwändig den Rest der Matrix (das heißt, die darunter liegenden Zeilen unter dem rechten oberen Element 204) berechnen oder bestimmen zu müssen. Alternativ kann auch der erste Grobschätzwert für die Tonart des Audiosignals auch aus dem zeitlich längsten Audiosignalsegment ermittelt werden, da dieses die für eine Auswertung der Tonartbestimmung höchste Informationsmenge der in dem Audiosignal enthaltenen Tonarten enthält. Diese längste Audiosignalsegment kann ein einfacher Form auch das zweite Audiosignalsegment ASS2 sein, da dieses zeitlich länger als das erste Audiosignalsegment ist.

- 14 -

In nachfolgenden Schritten kann dann jedoch die restliche Matrix aufgebaut werden und beispielsweise durch das vorstehend näher erläuterte Verfahren unter Verwendung einer Häufigkeitsauswertung die Tonart des Audiosignals präziser bestimmt werden, wobei dann zugleich einzelne Zeitsegmente in Bezug auf einen Tonartwechsel untersucht werden können.

Weiterhin zeigt sich im Bereich von 50 bis 90 Sekunden ein weiterer Tonartwechsel in der in Fig. 2A dargestellten Matrix 200. Dieser weitere Tonartwechsel zeigt jedoch keine derart "scharfen" Kanten wie die Kante 200, was mehrere Ursachen haben kann. Einerseits könnte beispielsweise ein Übergang zwischen den Tonarten gleitend sein (beispielsweise in der Form eines glissandos) und somit eine eindeutige oder abrupte Bestimmung eines Tonartwechsels unmöglich sein. Weiterhin könnte auch die Berücksichtigung der Zeitsegmente für die ersten 50 Sekunden des Audiosignals eine Rolle bei der Ermittlung des Tonartschätzwertes für die Audiosignalsegmente im Bereich um 70 Sekunden spielen, derart, dass eine scharfe Kante nicht mehr präzise ausgebildet werden kann, sondern vielmehr die den Tonartwechsel kennzeichnende Kante "Ausfransungen" aufweist. Weiterhin können auch Fehlklassifikationen auftreten, die beispielsweise durch lokale Disharmonien oder zeitlich zu kurze Zeitfenster bei glissandos ausgelöst werden können und somit eine eindeutige Tonarterkennung bzw. die Berechnung eines eindeutigen Tonartschätzwerts erschweren bzw. verhindern.

Weiterhin kann eine Information über die zweithäufigst auftretende Ausprägung eines Tonartschätzwerts zur Bestimmung der Tonart des Audiosignals verwendet werden. Hierbei ist insbesondere eine Paralleltonart zu einer Tonart zu berücksichtigen, wobei die Paralleltonart dadurch charakterisiert ist, dass diese die gleichen Töne in einer Oktave (wie die Grund-)Tonart aufweist. Beispielsweise wären in diesem Zusammenhang die Tonarten C-Dur und A-moll als Grund-Tonart und Paralleltonart zu nennen, da in beiden Tonarten die gleichen Töne in der Oktave

- 15 -

vorkommen, jedoch mit einer anderen Anordnung der Halbtonabstände.

5 Durch die Verwendung des zweithäufigsten Auftretens eines Tonartschätzwerts in Verbindung mit der Berücksichtigung mit der Paralleltonart für die Bestimmung der Tonart des Audiosignals ist es somit möglich, eine Verbesserung bzw. statistisch größere Absicherung des erfindungsgemäßen Ansatzes zur Bestimmung einer Tonart eines Audiosignals zu ermöglichen. Insbesondere
10 dann, wenn eine Grundtonart und eine Paralleltonart denjenigen Tonartschätzwerten entsprechen, die am häufigsten und am zweithäufigsten in der Matrix vorkommen, lässt sich mit einer deutlich höheren Sicherheit das Vorliegen entweder der (Grund)-Tonart oder der Paralleltonart im Audiosignal feststellen, als
15 dies lediglich durch die Berücksichtigung des am häufigsten vorkommenden Tonartschätzwerts in der Matrix möglich wäre.

Um eine weitere Verbesserung der Auswertung zu erreichen, kann ferner eine zeitliche Minimallänge eines Audiosignalsegments
20 von beispielsweise 2 Sekunden festgelegt werden. Hierdurch ist sichergestellt, dass beispielsweise bei dem Vorliegen des Audiosignals in zeitlich aufeinanderfolgenden Signalwerten eine Zeit-Frequenzbereichstransformation möglich ist, die eine zur Tonarterkennung ausreichende spektrale "Schärfe" aufweist, und
25 somit eine klare und eindeutige Trennung der einzelnen Töne in einer Oktave ermöglicht.

Zusammenfassend zu den Figuren 2A und 2B ist zu sagen, dass in diesem Ausführungsbeispiel eine Matrix aufgebaut wird, die An-
30 nahmen der Tonart über verschieden lange Zeitintervalle umfasst. Hierzu kann eine minimale Zeitfensterlänge (bzw. Zeitrahmenlänge) von beispielsweise 2 Sekunden definiert werden. Nachfolgend werden alle zusammenhängenden Bereiche untersucht, die von einem Startwert t_{start} bis zu einem Endwert t_{end} reichen,
35 wobei für die verschiedenen Bereiche zunächst die Endzeitpunkte t_{end} unterschiedlich sind und hieran anschließend ein neuer Startzeitpunkt t_{start} gewählt wird und wiederum Zeitfenster mit

- 16 -

unterschiedlichen Endzeitpunkten t_{end} mit dem neuen Startzeitpunkt t_{start} gewählt werden. Hierdurch werden benachbarte Zeitfenster oder Zeitrahmen generiert, die eine minimale Länge aufweisen und wobei sich die Zeitrahmen der gleichen Startzeitpunkte einander überlappen, jedoch mindestens in einem Zeitintervall von 2 Sekunden neue Informationen aus dem Audiosignal enthalten, die in vorhergehenden Segmenten noch nicht enthalten sind. Dies bedeutet somit, dass der erste Wert (Tonartschätzwert) für Noten der ersten 2 Sekunden, dann für die ersten vier Sekunden usw. geschätzt wird. Dieses Schätzen kann, wie vorstehend erläutert wurde, unter Verwendung von Tonartprofilen und einer Korrelation mit diesen Referenz-Tonartprofilen erfolgen, wobei dann die Ergebnisse aus der Korrelation, insbesondere der Tonartschätzwert der für das entsprechende Audiosignalsegment geschätzten Tonart in die einzelnen Elemente der ersten Zeile der Matrix beschrieben werden. Im nächsten Schritt wird der Startzeitpunkt ein Intervall von 2 Sekunden weiter (das heißt nach hinten) verschoben. In anderen Worten ausgedrückt wird die erste Berechnung dann das Zeitfenster zwischen zwei Sekunden und vier Sekunden, nachfolgend das Zeitfenster zwischen 2 und 6 Sekunden usw. betreffen. Die Ergebnisse für diese Abschätzung der Tonartschätzwerte wird dann in die zweite Zeile der sich ergebenden Matrix eingefügt. Die sich hierdurch ergebene obere Diagonalmatrix wird dann beispielsweise unter Verwendung einer Auswertung der Häufigkeitsverteilung der einzelnen Tonartschätzwerte untersucht und diese Tonartschätzwerte bzw. die Statistik oder Häufigkeitsverteilung der Tonartschätzwerte für das durch das Audiosignal mit den Musikinformationen gekennzeichnete Musikstück in einer Textdatei gespeichert.

Die mit dem vorliegenden Ansatz ermittelten Ergebnisse für die Bestimmung einer Tonart eines Audiosignals mit Musikinformationen ist in Fig. 3 dargestellt. Zur Ermittlung der in Fig. 3 dargestellten Ergebnisse wurden 12 Präludien und 12 Fugen von Johann Sebastian Bach's „the well-tempered clavier II“ untersucht, wobei gezeigt werden konnte, dass in 52,8% der Untersuchungen die richtige Tonart mit dem erfindungsgemäßen Ansatz

- 17 -

erkannt werden konnte. In 29,17% der Fälle wurde die Paralleltonart erkannt, wogegen in 10,42% die korrekte Tonart an zweiter Stelle in der Häufigkeitsverteilung aufgetreten ist. Lediglich in 8,33% wurde die falsche Tonart erkannt. Dies zeigt die hohe Leistungsfähigkeit des erfindungsgemäßen Ansatzes bei der Bestimmung einer Tonart eines Audiosignals mit Musikinformationen.

Abhängig von den Gegebenheiten kann das erfindungsgemäße Verfahren Bestimmen einer Tonart eines Audiosignals mit Musikinformationen in Hardware oder in Software implementiert werden. Die Implementierung kann auf einem digitalen Speichermedium, insbesondere einer Diskette oder CD mit elektronisch auslesbaren Steuersignalen erfolgen, die so mit einem programmierbaren Computersystem zusammenwirken können, dass das entsprechende Verfahren ausgeführt wird. Allgemein besteht die Erfindung somit auch in einem Computerprogrammprodukt mit einem auf einem maschinenlesbaren Träger gespeicherten Programmcode zur Durchführung des erfindungsgemäßen Verfahrens, wenn das Computerprogrammprodukt auf einem Rechner abläuft. Mit anderen Worten ausgedrückt, kann die Erfindung somit als ein Computerprogramm mit einem Programmcode zur Durchführung des Verfahrens realisiert werden, wenn das Computerprogramm auf einem Computer abläuft.

Zusammenfassend kann gesagt werden, dass im erfindungsgemäßen Ansatz ein Musikstück für verschiedene Zeitfenster untersucht wird. Die Länge der Fenster ergeben sich beispielsweise aus einem Vielfachen einer Mindestlänge von beispielsweise 250 Frames (= 250 Zeitrahmen). Zunächst wird ein Startpunkt auf 0 und ein Endpunkt auf 250 Zeitrahmen gesetzt. Für dieses Zeitfenster wird gezählt, welcher Halbton in wie vielen Zellen gesetzt ist. Die Ergebnisse werden beispielsweise in einem 12-zeiligen Vektor geschrieben. Dieser wird dann beispielsweise mit den Tonartprofilen korreliert und die Ergebnisse gespeichert. Der (12-zeilige) Vektor wird dann verschoben, sowie das aus dem zweiten Element das erste wird, aus dem ersten Element das zwölfte, usw.. Dieser Vektor und alle weiteren Umstellungen werden eben-

falls mit den Tonartprofilen korreliert. Eine Aussage darüber, welche Tonart für das betrachtete Zeitfenster angenommen werden kann, ergibt sich daraus, auf welchen Vektor und welches Tonartprofil das höchste Korrelationsergebnis erzielt wurde. Wird
5 beispielsweise für das zweite Profil und den dritten Vektor das höchste Ergebnis gefunden, handelt es sich um D-moll.

Anschließend wird der Endpunkt beispielsweise um weitere 250 Zeitrahmen (zeitlich) nach hinten verschoben und für dieses
10 Zeitfenster eine weitere Annahme gemacht. Dies wird wiederholt, bis der Endpunkt am Ende des Stücks angelangt ist. Dann wird der Startpunkt um 250 Zeitrahmen (zeitlich) nach hinten weitergesetzt und der Endpunkt beispielsweise auf 250
15 Zeitrahmen nach dem Startpunkt gesetzt, usw. . So werden für alle Zeitintervalle, die eine vielfache Länge von 250 Zeitrahmen besitzen, Annahmen für die Tonart gemacht. Alle diese Annahmen werden beispielsweise in eine Matrix als Zahlenwert (1 bis 24) eingetragen, wobei jeder Zahlenwert eine
20 entspricht. Anschließend werden die Annahmen für jede Tonart gezählt und die Ergebnisse in Prozent in eine Textdatei ausgegeben. Unter Berücksichtigung dieser Textdatei kann nun entweder direkt eine Tonart des gesamten Musikstücks ermittelt werden oder aber ein Wechsel einer Tonart innerhalb des
25 Musikstücks bestimmt werden. Alternativ kann auch die Bestimmung der Tonart des Musikstücks und ein Zeitpunkt von Tonartwechsel bzw. die Tonarten zwischen denen gewechselt wird, durch die Auswertung der Matrix bestimmt werden.

Patentansprüche

1. Vorrichtung (100) zum Bestimmen einer Tonart eines Audio-
signals (108) mit Musikinformationen, die folgende Merkma-
le aufweist:

eine Einrichtung (102) zum Definieren eines ersten Audio-
signalsegments (ASS1) und eines zweiten Audiosignalseg-
ments (ASS2), wobei sich das erste Audiosignalsegment
(ASS1) und das zweite Audiosignalsegment (ASS2) überlappen
und unterschiedliche Zeitdauern haben;

eine Einrichtung (104) zum Abschätzen einer Tonart für je-
des Audiosignalsegment (ASS1, ASS2), um einen ersten Ton-
artschätzwert für das erste Audiosignalsegment (ASS1) und
einen zweiten Tonartschätzwert für das zweite Audiosignal-
segment (ASS2) zu erhalten; und

eine Einrichtung (106) zum Auswerten der Tonartschätzwerte
(112), um auf der Basis des ersten Tonartschätzwerts und
des zweiten Tonartschätzwerts die Tonart (114) des Audio-
signals (108) zu bestimmen.

2. Vorrichtung (100) zum Bestimmen gemäß Anspruch 1, bei der
die Einrichtung (102) zum Definieren ausgebildet ist, um
das erste Audiosignalsegment (ASS1) in einer ersten Zeit-
spanne zwischen einem Startzeitpunkt und einem ersten End-
zeitpunkt zu definieren und das zweite Audiosignalsegment
(ASS2) in einer zweiten Zeitspanne zwischen dem Startzeit-
punkt und einem zweiten Endzeitpunkt zu definieren, wobei
der zweite Endzeitpunkt zeitlich dem ersten Endzeitpunkt
nachgelagert ist.

3. Vorrichtung (100) zum Bestimmen gemäß Anspruch 1 oder 2,
bei der die Einrichtung (102) zum Definieren ferner ausge-
bildet ist, um ein drittes Audiosignalsegment (ASS3) in

einer dritten Zeitspanne zwischen einem zweiten Startzeitpunkt und einem dritten Endzeitpunkt zu definieren, wobei der zweite Startzeitpunkt dem Startzeitpunkt zeitlich nachgelagert ist, wobei die Einrichtung (104) zum Abschätzen ausgebildet ist, um einen dritten Tonartschätzwert für das dritte Audiosignalsegment (ASS3) zu erhalten und wobei die Einrichtung (106) zum Auswerten ausgebildet ist, um den dritten Tonartschätzwert für das Bestimmen der Tonart (114) des Audiosignals (108) zu verwenden.

4. Vorrichtung (100) zum Bestimmen gemäß Anspruch 3, bei der die Einrichtung (102) zum Definieren ausgebildet ist, um ein viertes Audiosignalsegment (ASS4) in einer vierten Zeitspanne zwischen dem zweiten Startzeitpunkt und einem vierten Endzeitpunkt zu definieren, wobei der vierte Endzeitpunkt zeitlich dem dritten Endzeitpunkt nachgelagert ist, wobei die Einrichtung (104) zum Abschätzen ausgebildet ist, um einen vierten Tonartschätzwert für das vierte Audiosignalsegment (ASS4) zu erhalten und wobei die Einrichtung (106) zum Auswerten ausgebildet ist, um den vierten Tonartschätzwert für das Bestimmen der Tonart (114) des Audiosignals (108) zu verwenden.
5. Vorrichtung (100) zum Bestimmen gemäß einem der Ansprüche 3 oder 4, bei der die Einrichtung (106) zum Auswerten ausgebildet ist, um eine Erstschätzung der Tonart (114) des Audiosignals (108) auf der Grundlage des zweiten Tonartschätzwertes zu bestimmen und dann die Tonart (114) des Audiosignals (108) auf der Basis des ersten Tonartschätzwertes und des zweiten Tonartschätzwertes zu bestimmen.
6. Vorrichtung (100) zum Bestimmen gemäß Anspruch 5, bei der die Einrichtung (106) ausgebildet ist, um eine Häufigkeit eines Auftretens einer Ausprägung für die erhaltenen Tonartschätzwerte (112) zu ermitteln und die Tonart (114) des

Audiosignals (108) auf der Basis einer häufigsten aufgetretenen Ausprägung eines Tonartschätzwertes zu bestimmen.

- 5 7. Vorrichtung (100) zum Bestimmen gemäß Anspruch 6, bei der die Einrichtung (106) zum Auswerten ausgebildet ist, um eine zweit-häufigste aufgetretene Ausprägung eines Tonartschätzwertes oder einen Tonartschätzwert für eine Paralleltonart zu der den häufigsten aufgetretenen Tonartschätzwert entsprechenden Tonart für das Bestimmen der
10 Tonart (114) des Audiosignals (108) zu bestimmen.
8. Vorrichtung (100) zum Bestimmen gemäß einem der Ansprüche 1 bis 7, bei der die Einrichtung (102) zum Definieren ausgebildet ist, um die Audiosignalsegmente (ASS1, ASS2) der-
15 art zu definieren, dass sie mindestens eine vorbestimmte zeitliche Länge aufweisen.
9. Vorrichtung (100) zum Bestimmen gemäß einem der Ansprüche 3 bis 8, bei der die Einrichtung (104) zum Abschätzen ausgebildet ist, um die für die Audiosignalsegmente (ASS1 bis
20 ASS3) erhaltenen Tonartschätzwerte in Elemente einer Matrix einzutragen, wobei diejenigen Elemente einer Zeile der Matrix Audiosignalsegmenten (ASS1, ASS2) entsprechen, die einen gemeinsamen Startzeitpunkt (t_{start}) aufweisen und
25 diejenigen Elemente einer Spalte der Matrix Audiosignalsegmenten (ASS2, ASS3) entsprechen, die einen gemeinsamen Endzeitpunkt (t_{end}) aufweisen, wobei die Einrichtung (106) zum Auswerten ferner eine Einrichtung zum Ermitteln eines Tonartwechsels aufweist, die ausgebildet ist, um als Zeit-
30 punkt eines Tonartwechsels denjenigen Start- (t_{start}) oder Endzeitpunkt (t_{end}) zu ermitteln, bei dem ein in ein Referenzelement der Matrix eingetragener Referenz-Tonartschätzwert von einem Test-Tonartschätzwert verschieden ist, der in einem zu dem Referenzelement in der Matrix
35 benachbart angeordneten Testelement der Matrix eingetragen ist.

10. Vorrichtung (100) zum Bestimmen gemäß Anspruch 9, bei der die Tonartschätzwerte voneinander unterscheidbare Zahlen sind und bei der die Einrichtung zum Ermitteln ausgebildet ist, um einen Tonartwechsel dann zu ermitteln, wenn eine Differenz zwischen dem Referenztonartschätzwert und dem Test-Tonartschätzwert größer als ein vordefinierter Schwellwert ist.
11. Vorrichtung (100) zum Bestimmen gemäß Anspruch 10, bei der die Einrichtung zum Ermitteln ausgebildet ist, um einen Tonartwechsel dann festzustellen, wenn das Referenzelement und das Testelement in einer Zeile der Matrix angeordnet sind und keinen Tonartwechsel festzustellen, wenn das Referenzelement und das Testelement in einer Spalte der Matrix angeordnet sind.
12. Verfahren zum Bestimmen einer Tonart eines Audiosignals (108) mit Musikinformationen, das folgende Schritte aufweist:
- Definieren eines ersten Audiosignalsegments (ASS1) und eines zweiten Audiosignalsegments (ASS2), wobei sich das erste Audiosignalsegment (ASS1) und das zweite Audiosignalsegment (ASS2) überlappen und unterschiedliche Zeitdauern haben;
- Abschätzen einer Tonart für jedes Audiosignalsegment (ASS1, ASS2), um einen ersten Tonartschätzwert für das erste Audiosignalsegment (ASS1) und einen zweiten Tonartschätzwert für das zweite Audiosignalsegment (ASS2) zu erhalten; und
- Auswerten der Tonartschätzwerte (112), um auf der Basis des ersten Tonartschätzwerts und des zweiten Tonartschätz-

- 23 -

werts die Tonart (114) des Audiosignals (108) zu bestimmen.

13. Computerprogramm zur Durchführung des Verfahrens gemäß Anspruch 12, wenn das Programm auf einem Computer abläuft.
- 5

FIG. 1

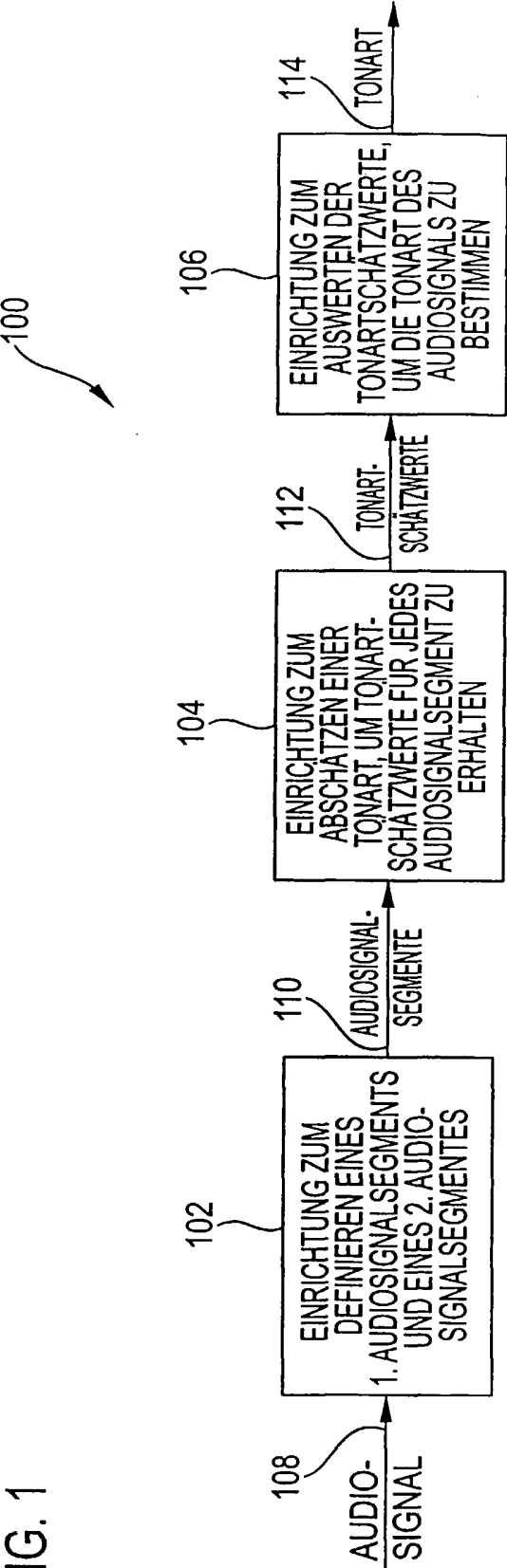
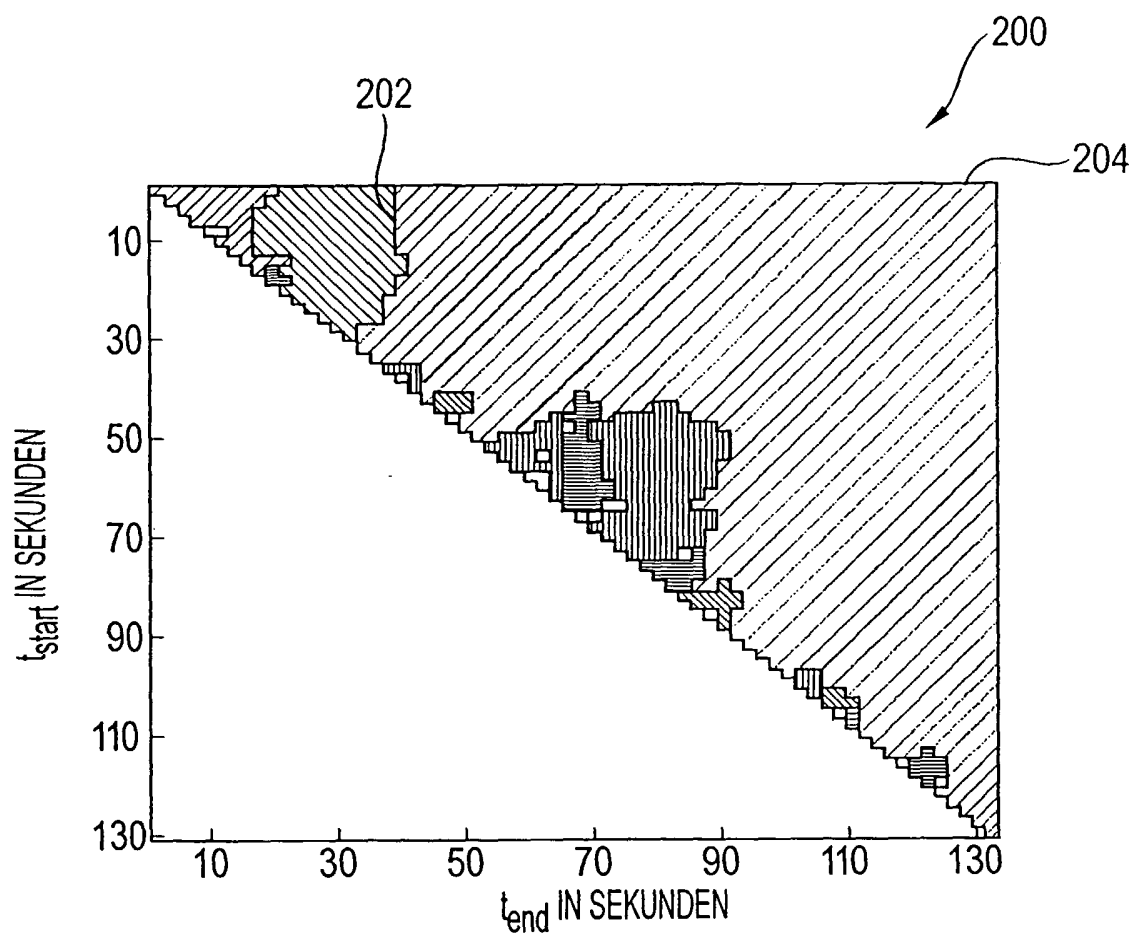
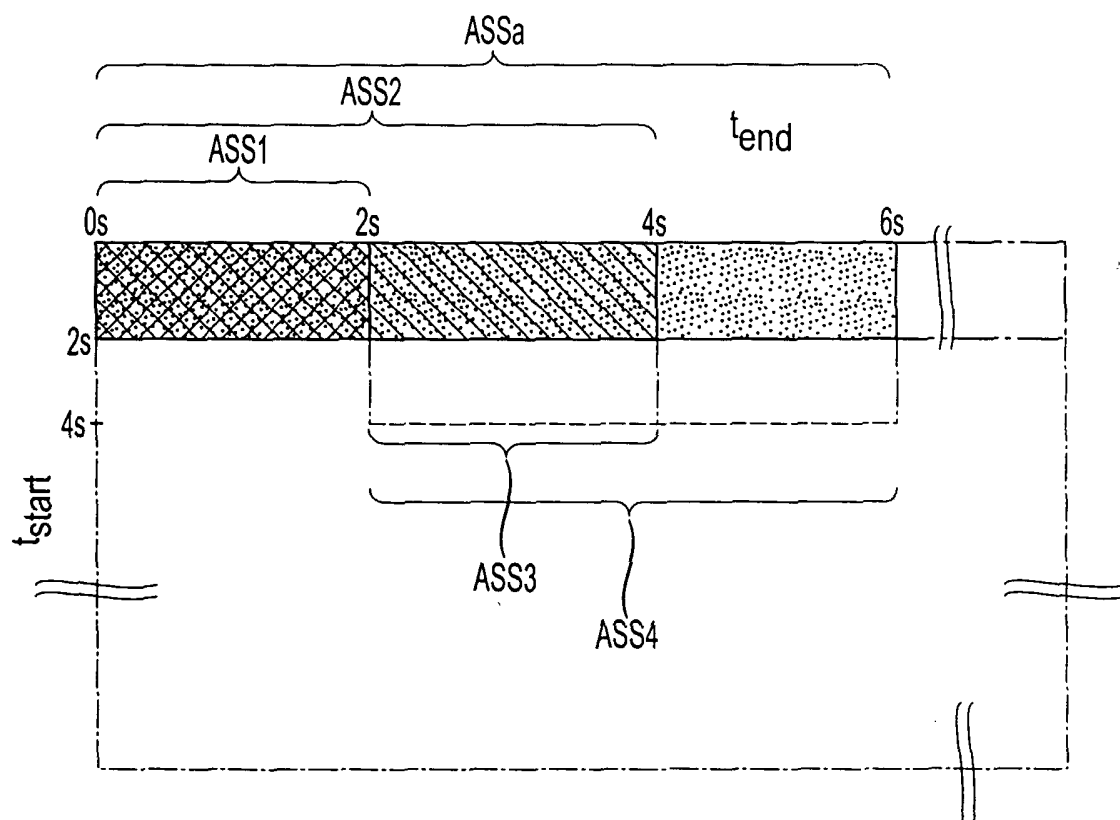


FIG. 2A



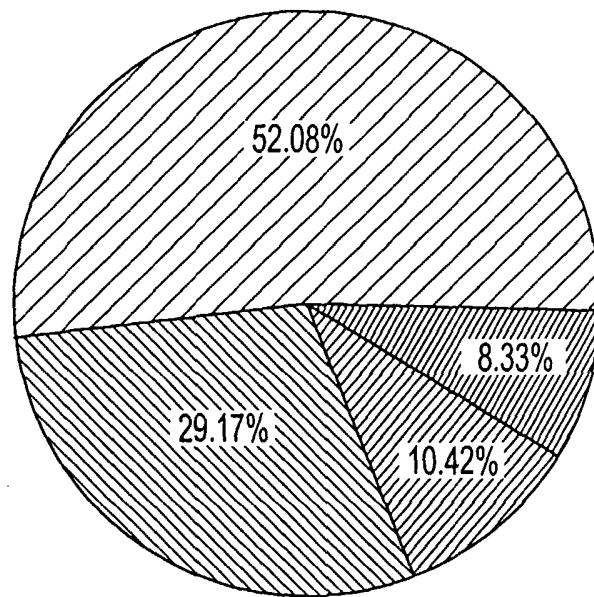
	A#-MOLL		A#-DUR
	G#-MOLL		G#-DUR
	F#-MOLL		F#-DUR
	D#-MOLL		D#-DUR
	B#-DUR		C#-DUR

FIG. 2B



4/4

FIG. 3



KORREKTE TONART



PARALLELTONART



KORREKTE TONART AN ZWEITER STELLE



FALSCHER TONART

INTERNATIONAL SEARCH REPORT

 International Application No
 PCT/EP2005/002304

 A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 G10H1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G10H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	IZMIRLI O ET AL: "Recognition of musical tonality from sound input" ELECTROTECHNICAL CONFERENCE, 1994. PROCEEDINGS., 7TH MEDITERRANEAN ANTALYA, TURKEY 12-14 APRIL 1994, NEW YORK, NY, USA, IEEE, 12 April 1994 (1994-04-12), pages 269-271, XP010130828 ISBN: 0-7803-1772-6 abstract -----	1-4, 8, 12, 13
A	US 5 756 918 A (FUNAKI ET AL) 26 May 1998 (1998-05-26) abstract; figures 2, 4a-c, 5a-b, 6a-b, 16, 17, 18a-b, 19a-b, 20, 21 column 2, line 40 - line 65 column 15, line 19 - column 18, line 60 tables 7-13 ----- -/--	1-13

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

24 June 2005

Date of mailing of the international search report

26/08/2005

Name and mailing address of the ISA

 European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer

Lecoainte, M

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP2005/002304

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BARTSCH M A ET AL: "To catch a chorus: using chroma-based representations for audio thumbnailing" APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 2001 IEEE WORKSHOP ON THE OCT. 21-24, 2001, PISCATAWAY, NJ, USA,IEEE, 21 October 2001 (2001-10-21), pages 15-18, XP010566863 ISBN: 0-7803-7126-7 abstract 2. The Chromagram and Chroma-Based Feature Class 3. Algorithm Description figures 1-4 -----	9-11

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP2005/002304

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
US 5756918	A	26-05-1998	JP	3307156 B2	24-07-2002
			JP	8292762 A	05-11-1996

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP2005/002304

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

IPK 7 G10H1/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 7 G10H

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, INSPEC, WPI Data

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	IZMIRLI O ET AL: "Recognition of musical tonality from sound input" ELECTROTECHNICAL CONFERENCE, 1994. PROCEEDINGS., 7TH MEDITERRANEAN ANTALYA, TURKEY 12-14 APRIL 1994, NEW YORK, NY, USA, IEEE, 12. April 1994 (1994-04-12), Seiten 269-271, XP010130828 ISBN: 0-7803-1772-6 Zusammenfassung -----	1-4, 8, 12, 13
A	US 5 756 918 A (FUNAKI ET AL) 26. Mai 1998 (1998-05-26) Zusammenfassung; Abbildungen 2, 4a-c, 5a-b, 6a-b, 16, 17, 18a-b, 19a-b, 20, 21 Spalte 2, Zeile 40 - Zeile 65 Spalte 15, Zeile 19 - Spalte 18, Zeile 60 Tabellen 7-13 ----- -/-	1-13

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

A Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

E älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

L Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

O Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

P Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

T Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

X Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

Y Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

G Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

24. Juni 2005

Absenddatum des internationalen Recherchenberichts

26/08/2005

Name und Postanschrift der Internationalen Recherchenbehörde
Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Lecointe, M

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	BARTSCH M A ET AL: "To catch a chorus: using chroma-based representations for audio thumbnailing" APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 2001 IEEE WORKSHOP ON THE OCT. 21-24, 2001, PISCATAWAY, NJ, USA, IEEE, 21. Oktober 2001 (2001-10-21), Seiten 15-18, XP010566863 ISBN: 0-7803-7126-7 Zusammenfassung 2. The Chromagram and Chroma-Based Feature Class 3. Algorithm Description Abbildungen 1-4 -----	9-11

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

PT/EP2005/002304

Formblatt PCT/ISA/210 (Anhang Patentfamilie) (Januar 2004)