



(11) **EP 1 794 743 B1**

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
24.04.2013 Patentblatt 2013/17

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

(21) Anmeldenummer: **05760763.2**

(86) Internationale Anmeldenummer:
PCT/EP2005/007751

(22) Anmeldetag: **15.07.2005**

(87) Internationale Veröffentlichungsnummer:
WO 2006/034743 (06.04.2006 Gazette 2006/14)

(54) **VORRICHTUNG UND VERFAHREN ZUM GRUPPIEREN VON ZEITLICHEN SEGMENTEN EINES MUSIKSTÜCKS**

DEVICE AND METHOD FOR ARRANGING IN GROUPS TEMPORAL SEGMENTS OF A PIECE OF MUSIC

DISPOSITIF ET PROCEDE POUR REGROUPER DES SEGMENTS TEMPORELS D'UN MORCEAU DE MUSIQUE

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

(30) Priorität: **28.09.2004 DE 102004047068**

(43) Veröffentlichungstag der Anmeldung:
13.06.2007 Patentblatt 2007/24

(73) Patentinhaber: **Sony Corporation**
Tokyo 108-0075 (JP)

(72) Erfinder:
• **VAN PINXTEREN, Markus**
80469 München (DE)
• **SAUPE, Michael**
98693 Ilmenau (DE)
• **CREMER, Markus**
Berkeley, CA 94707 (US)

(74) Vertreter: **Lewis, Darren John et al**
D Young & Co LLP
120 Holborn
London EC1N 2DY (GB)

(56) Entgegenhaltungen:
WO-A-2004/038694 US-B1- 6 542 869

- **ROGER DANNENBERG AND NING HU:**
"Discovering Musical Structure in Audio Recordings" PROC. 2ND INT. CONFERENCE IN MUSIC & ARTIFICIAL INTELLIGENCE (ICMAI), 2002, XP002348414
- **LOGAN B ET AL: "Music summarization using key phrases" ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 2000. ICASSP '00. PROCEEDINGS. 2000 IEEE INTERNATIONAL CONFERENCE ON 5-9 JUNE 2000, PISCATAWAY, NJ, USA, IEEE, Bd. 2, 5. Juni 2000 (2000-06-05), Seiten 749-752, XP010504831 ISBN: 0-7803-6293-4**
- **FOOTE J T ET AL: "MEDIA SEGMENTATION USING SELF-SIMILARITY DECOMPOSITION" PROCEEDINGS OF THE SPIE, SPIE, BELLINGHAM, VA, US, Bd. 5021, 2003, Seiten 167-175, XP009052550 ISSN: 0277-786X**

Anmerkung: Innerhalb von neun Monaten nach Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents im Europäischen Patentblatt kann jedermann nach Maßgabe der Ausführungsordnung beim Europäischen Patentamt gegen dieses Patent Einspruch einlegen. Der Einspruch gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist. (Art. 99(1) Europäisches Patentübereinkommen).

EP 1 794 743 B1

Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf die Audiosegmentierung und insbesondere auf die Analyse von Musikstücken auf die in den Musikstücken enthaltenen einzelnen Hauptteile, die in dem Musikstück wiederholt auftreten können.

[0002] Musik aus dem Rock- und Popbereich besteht meistens aus mehr oder weniger eindeutigen Segmenten, wie beispielsweise Intro, Strophe, Refrain, Bridge, Outro, etc. Die Anfangs- und Endzeitpunkte solcher Segmente zu detektieren und die Segmente nach ihrer Zugehörigkeit zu den wichtigsten Klassen (Strophe und Refrain) zu gruppieren, ist Ziel der Audiosegmentierung. Eine korrekte Segmentierung und auch Kennzeichnung der berechneten Segmente kann in verschiedenen Bereichen sinnvoll eingesetzt werden. Beispielsweise können so Musikstücke von Online-Anbietern, wie Amazon, Musicline, etc. intelligent "angespielt" werden.

[0003] Die meisten Anbieter im Internet beschränken sich bei ihren Hörbeispielen auf einen kurzen Ausschnitt aus den angebotenen Musikstücken. In diesem Fall wäre es natürlich auch sinnvoll, dem Interessenten nicht nur die ersten 30 Sekunden oder beliebige 30 Sekunden, sondern einen möglichst repräsentativen Ausschnitt aus dem Lied anzubieten. Dies könnte z. B. der Refrain sein, oder aber auch eine Zusammenfassung des Liedes, bestehend aus Segmenten, die den verschiedenen Hauptklassen (Strophe, Refrain, ...) angehören.

[0004] Ein weiteres Anwendungsbeispiel für die Technik der Audiosegmentierung ist das Integrieren des Segmentierungs-/Gruppierungs-/Markierungsalgorithmus in einen Musicplayer. Die Informationen über Segmentanfänge und Segmentenden ermöglichen das gezielte Navigieren durch ein Musikstück. Durch die Klassenzugehörigkeit der Segmente, also ob ein Segment eine Strophe, ein Refrain, etc. ist, kann z. B. auch direkt zum nächsten Refrain oder zur nächsten Strophe gesprungen werden. Eine derartige Anwendung ist für große Musikmärkte von Interesse, die ihren Kunden die Möglichkeit bieten, in komplette Alben hinein zu hören. Dadurch erspart sich der Kunde das lästige, suchende Vorspulen zu charakteristischen Stellen im Lied, die ihn vielleicht dazu bewegen könnten, ein Musikstück am Ende tatsächlich zu kaufen.

[0005] Auf dem Gebiet der Audiosegmentierung existieren verschiedene Ansätze. Nachfolgend wird der Ansatz von Jonathan Foote und Matthew Cooper beispielhaft dargestellt. Dieses Verfahren ist in FOOTE, J.T. / Cooper, M.L.: Summarizing Popular Music via Structural Similarity Analysis. Proceedings of the IEEE Workshop of Signal Processing to Audio and Acoustics 2003. FOOTE, J.T. / COOPER, M.L.: Media Segmentation using Self-Similar Decomposition. Proceedings of SPIE Storage and Retrieval for Multimedia Databases, Bd. 5021, S. 167-75, Januar 2003, dargestellt.

[0006] Das bekannte Verfahren von Foote wird anhand des Blockschaltbilds von Fig. 5 beispielhaft erläutert. Zunächst wird eine WAV-Datei 500 bereitgestellt. In einem nachgeschalteten Extraktionsblock 502 findet dann eine Merkmalsextraktion statt, wobei als Merkmal die Spektralkoeffizienten an sich oder alternativ die Mel-Frequenz-Cepstral-Koeffizienten (MFCCs) extrahiert werden. Vor dieser Extraktion wird eine Kurzzeit-Fourier-Transformation (STFT) mit 0,05 Sekunden breiten nicht-überlappenden Fenstern mit der WAV-Datei durchgeführt. Die MFCC-Merkmale werden dann im Spektralbereich extrahiert. Hierbei sei darauf hingewiesen, dass die Parametrisierung nicht für eine Kompression, Übertragung oder Rekonstruktion optimiert ist, sondern für eine Audioanalyse. Die Anforderung besteht dahingehend, dass ähnliche Audiostücke ähnliche Merkmale erzeugen.

[0007] Die extrahierten Merkmale werden dann in einem Speicher 504 abgelegt.

[0008] Auf den Merkmalsextraktionsalgorithmus findet nunmehr ein Segmentierungsalgorithmus statt, der in einer Ähnlichkeitsmatrix endet, wie es in einem Block 506 dargestellt ist. Zunächst wird jedoch die Merkmalsmatrix eingelesen (508), um dann Merkmalsvektoren zu gruppieren (510), um dann auf Grund der gruppierten Merkmalsvektoren eine Ähnlichkeitsmatrix aufzubauen, die aus einer Distanzmessung zwischen jeweils allen Merkmalen besteht. Im Einzelnen werden alle paarweisen Kombinationen von Audiofenstern unter Verwendung eines quantitativen Ähnlichkeitsmaßes, also der Distanz, verglichen.

[0009] Der Aufbau der Ähnlichkeitsmatrix ist in Fig. 8 dargestellt. So ist in Fig. 8 das Musikstück als Strom oder Stream 800 von Audioabstastwerten dargestellt. Das Audiostück wird, wie es ausgeführt worden ist, gefenstert, wobei ein erstes Fenster mit i und ein zweites Fenster mit j bezeichnet sind. Insgesamt hat das Audiostück z. B. K Fenster. Dies bedeutet, dass die Ähnlichkeitsmatrix K Zeilen und K Spalten hat. Dann wird für jedes Fenster i und für jedes Fenster j ein Ähnlichkeitsmaß zueinander berechnet, wobei das berechnete Ähnlichkeitsmaß oder Distanzmaß $D(i,j)$ an der durch i und j bezeichneten Zeile bzw. Spalte in der Ähnlichkeitsmatrix eingegeben wird. Eine Spalte zeigt daher die Ähnlichkeit des durch j bezeichneten Fensters zu allen anderen Audiofenstern in dem Musikstück. Die Ähnlichkeit des Fensters j zum allerersten Fenster des Musikstücks würde dann in der Spalte j und in der Zeile 1 stehen. Die Ähnlichkeit des Fensters j zum zweiten Fenster des Musikstücks würde dann in der Spalte j, aber nunmehr in der Zeile 2 stehen. Dagegen würde die Ähnlichkeit des zweiten Fensters zum ersten Fenster in der zweiten Spalte der Matrix und in der ersten Zeile der Matrix stehen.

[0010] Es ist zu sehen, dass die Matrix dahingehend redundant ist, dass sie zur Diagonalen symmetrisch ist, und dass auf der Diagonalen die Ähnlichkeit eines Fensters zu sich selbst steht, was den trivialen Fall einer 100%-igen Ähnlichkeit darstellt.

[0011] Ein Beispiel für eine Ähnlichkeitsmatrix eines Stücks ist in Fig. 6 zu sehen. Hier ist wieder die komplett symmetrische Struktur der Matrix bezüglich der Hauptdiagonalen erkennbar, wobei die Hauptdiagonale als heller Streifen ersichtlich ist. Ferner wird darauf hingewiesen, dass auf Grund der kleinen Fensterlänge im Vergleich zu der relativ groben Zeitauflösung in Fig. 6 die Hauptdiagonale nicht als heller durchgehender Strich zu sehen ist, sondern aus Fig. 6 nur in etwa erkennbar ist.

[0012] Hierauf wird unter Verwendung der Ähnlichkeitsmatrix, wie sie z. B. in Fig. 6 dargestellt ist, eine Kernel-Korrelation 512 mit einer Kernel-Matrix 514 durchgeführt, um ein Neuheitsmaß, das auch als "Novelty Score" bekannt ist, zu erhalten, das gemittelt werden könnte und in geglätteter Form in Fig. 9 dargestellt ist. Die Glättung dieses Novelty Scores ist in Fig. 5 schematisch durch einen Block 516 dargestellt.

[0013] Hierauf werden in einem Block 518 die Segmentgrenzen unter Verwendung des geglätteten Neuheitswertverlaufs ausgelesen, wobei hierzu die lokalen Maxima im geglätteten Neuheitsverlauf ermittelt und gegebenenfalls noch um eine durch die Glättung verursachte konstante Anzahl von Samples verschoben werden müssen, um tatsächlich die richtigen Segmentgrenzen des Audiostreams als absolute oder relative Zeitangabe zu erhalten.

[0014] Hierauf wird, wie es bereits in einem mit Clustering bezeichneten Block aus Fig. 5 ersichtlich ist, eine so genannte Segmentähnlichkeitsdarstellung oder Segmentähnlichkeitsmatrix erstellt. Ein Beispiel für eine Segmentähnlichkeitsmatrix ist in Fig. 7 dargestellt. Die Ähnlichkeitsmatrix in Fig. 7 ist prinzipiell ähnlich zu der Merkmals-Ähnlichkeitsmatrix von Fig. 6, wobei nun jedoch nicht mehr, wie in Fig. 6, Merkmale aus Fenstern verwendet werden, sondern Merkmale aus einem ganzen Segment. Die Segmentähnlichkeitsmatrix hat eine ähnliche Aussage wie die Merkmals-ähnlichkeitsmatrix, jedoch mit einer wesentlich größeren Auflösung, was natürlich gewünscht ist, wenn betrachtet wird, dass Fensterlängen im Bereich von 0,05 Sekunden liegen, während vernünftig lange Segmente im Bereich von vielleicht 10 Sekunden eines Stückes liegen.

[0015] Hierauf wird dann in einem Block 522 ein Clustering durchgeführt, also eine Einordnung der Segmente in Segmentklassen (eine Einordnung von ähnlichen Segmenten in dieselbe Segmentklasse), um dann in einem Block 524 die gefundenen Segmentklassen zu markieren, was auch als "Labeling" bezeichnet wird. So wird im Labeling ermittelt, welche Segmentklasse Segmente enthält, die Strophen sind, die Refrains sind, die Intros, Outros, Bridges, etc. sind.

[0016] Schließlich wird in einem mit 526 in Fig. 5 bezeichneten Block eine Musicsummary erstellt, die z. B. einem Benutzer bereitgestellt werden kann, um ohne Redundanz von einem Stück nur z. B. eine Strophe, einen Refrain und das Intro zu hören.

[0017] Nachfolgend wird auf die einzelnen Blöcke noch detaillierter eingegangen.

[0018] Wie es bereits ausgeführt worden ist, findet die eigentliche Segmentierung des Musikstücks erst dann statt, wenn die Merkmalsmatrizen generiert und gespeichert sind (Block 504).

[0019] Je nach dem, anhand welchen Merkmals das Musikstück auf seine Struktur hin untersucht werden soll, wird die entsprechende Merkmalsmatrix ausgelesen und zur Weiterverarbeitung in einen Arbeitsspeicher geladen. Die Merkmalsmatrix hat die Dimension Anzahl der Analysefenster mal Anzahl der Merkmalskoeffizienten.

[0020] Durch die Ähnlichkeitsmatrix wird der Merkmalsverlauf eines Stücks in eine zweidimensionale Darstellung gebracht. Für jede paarweise Kombination von Merkmalsvektoren wird das Distanzmaß berechnet, das in der Ähnlichkeitsmatrix festgehalten wird. Zur Berechnung des Distanzmaßes zwischen zwei Vektoren gibt es verschiedene Möglichkeiten, nämlich beispielsweise die Euklidische Distanzmessung und die Cosinus-Distanzmessung. Ein Ergebnis $D(i,j)$ zwischen den zwei Merkmalsvektoren wird im i,j -ten Element der Fenster-Ähnlichkeitsmatrix (Block 506) gespeichert. Die Hauptdiagonale der Ähnlichkeitsmatrix repräsentiert den Verlauf über das gesamte Stück. Dementsprechend resultieren die Elemente der Hauptdiagonalen aus dem jeweiligen Vergleich eines Fensters mit sich selbst und weisen immer den Wert der größten Ähnlichkeit auf. Bei der Cosinus-Distanzmessung ist dies der Wert 1, bei der einfachen skalaren Differenz und der Euklidischen Distanz ist dieser Wert gleich 0.

[0021] Zur Visualisierung einer Ähnlichkeitsmatrix, wie sie in Fig. 6 dargestellt ist, bekommt jedes Element i, j einen Grauwert zugewiesen. Die Grauwerte sind proportional zu den Ähnlichkeitswerten abgestuft, so dass die maximale Ähnlichkeit (die Hauptdiagonale) der maximalen Ähnlichkeit entspricht. Durch diese Darstellung kann man die Struktur eines Liedes bereits optisch auf Grund der Matrix erkennen. Bereiche ähnlicher Merkmalsausprägung entsprechen Quadranten ähnlicher Helligkeit entlang der Hauptdiagonalen. Die Grenzen zwischen den Bereichen zu finden, ist die Aufgabe der eigentlichen Segmentierung.

[0022] Die Struktur der Ähnlichkeitsmatrix ist wichtig für das in der Kernel-Korrelation 512 berechnete Neuheitsmaß. Das Neuheitsmaß entsteht durch die Korrelation eines speziellen Kerns entlang der Hauptdiagonalen der Ähnlichkeitsmatrix. Ein beispielhafter Kernel K ist in Fig. 5 dargestellt. Korreliert man diese Kernel-Matrix entlang der Hauptdiagonalen der Ähnlichkeitsmatrix S , und summiert dabei alle Produkte der übereinander liegenden Matrixelemente für jeden Zeitpunkt i des Stücks, so erhält man das Neuheitsmaß, das in geglätteter Form beispielhaft in Fig. 9 dargestellt ist. Vorzugsweise wird nicht der Kernel K in Fig. 5 verwendet, sondern ein vergrößerter Kernel, der zudem mit einer Gaussverteilung überlagert ist, so dass die Kanten der Matrix gegen 0 streben.

[0023] Die Selektion der markanten Maxima im Neuheitsverlauf ist wichtig für die Segmentierung. Die Auswahl aller Maxima des ungeglätteten Neuheitsverlaufs würde zu einer starken Übersegmentierung des Audiosignals führen.

[0024] Daher sollte das Neuheitsmaß geglättet werden, und zwar mit verschiedenen Filtern, wie beispielsweise IIR-Filter oder FIR-Filter.

[0025] Sind die Segmentgrenzen eines Musikstücks extrahiert, so müssen nun ähnliche Segmente als solche gekennzeichnet und in Klassen gruppiert werden.

[0026] Foote und Cooper beschreiben die Berechnung einer segmentbasierten Ähnlichkeitsmatrix mittels einer Cullback-Leibler-Distanz. Hierfür werden anhand der aus dem Neuheitsverlauf gewonnenen Segmentgrenzen einzelne Segmentmerkmalsmatrizen aus der gesamten Merkmalsmatrix extrahiert, d. h. jede dieser Matrizen ist eine Submatrix der gesamten Merkmalsmatrix. Die damit entstandene Segmentähnlichkeitsmatrix 520 wird nunmehr einer Singularwertzerlegung (SVD; SVD = Singular Value Decomposition) unterzogen. Hierauf erhält man Singulärwerte in absteigender Reihenfolge.

[0027] Im Block 526 wird dann eine automatische Zusammenfassung eines Stücks anhand der Segmente und Cluster eines Musikstücks durchgeführt. Hierzu werden zunächst die zwei Cluster mit den größten Singulärwerten ausgewählt. Dann wird das Segment mit dem Maximalwert des entsprechenden Cluster-Indikators zu dieser Summary hinzugefügt. Dies bedeutet, dass die Summary eine Strophe und einen Refrain umfasst. Alternativ können auch alle wiederholten Segmente entfernt werden, um sicherzustellen, dass alle Informationen des Stücks bereitgestellt werden, jedoch immer genau einmal.

[0028] Bezüglich weiterer Techniken zur Segmentierung/Musikanalyse wird auf CHU, s. / LOGAN B.: Music Summary using Key Phrases. Technical Report, Cambridge Research Laboratory 2000, BARTSCH, M.A. / WAKEFIELD, g. H.: To Catch a Chorus: Using Chroma-Based Representation for Audio Thumbnailing. Proceedings of the IEEE Workshop of Signal Processing to Audio and Acoustics 2001. http://musen.engin.umich.edu/papers/bartsch_wakefield_waspaa01_final.pdf, verwiesen

[0029] Nachteilig an dem bekannten Verfahren ist die Tatsache, dass die Singularwertzerlegung (SVD) zur Segmentklassenbildung, also zum Zuordnen von Segmenten zu Clustern zum einen sehr rechenaufwendig ist und zum anderen problematisch in der Beurteilung der Ergebnisse ist. So wird dann, wenn die Singulärwerte nahezu gleich groß sind, eine möglicherweise falsche Entscheidung dahingehend getroffen, dass die beiden ähnlichen Singulärwerte eigentlich die gleiche Segmentklasse und nicht zwei unterschiedliche Segmentklassen darstellen.

[0030] Ferner wurde herausgefunden, dass die Ergebnisse, die durch die Singularwertzerlegung erhalten werden, dann immer problematischer werden, wenn es starke Ähnlichkeitswertunterschiede gibt, wenn also ein Stück sehr ähnliche Anteile enthält, wie Strophe und Refrain, aber auch relativ unähnliche Anteile, wie Intro, Outro oder Bridge.

[0031] Ferner problematisch an dem bekannten Verfahren ist, dass immer davon ausgegangen wird, dass der Cluster unter den beiden Clustern mit den höchsten Singulärwerten, der das erste Segment im Lied hat, der Cluster "Strophe" ist, und dass der andere Cluster der Cluster "Refrain" ist. Dieses Prozedere basiert darauf, dass man im bekannten Verfahren davon ausgeht, dass ein Lied immer mit einer Strophe beginnt. Erfahrungen haben gezeigt, dass damit erhebliche Labelingfehler erhalten werden. Dies ist insofern problematisch, da das Labeling gewissermaßen die "Ernte" des gesamten Verfahrens ist, also das, was der Benutzer unmittelbar erfährt. Waren die vorhergehenden Schritte noch so präzise und aufwändig, so relativiert sich alles, wenn am Ende falsch gelabelt wird, da dann beim Benutzer das Vertrauen in das gesamte Konzept insgesamt Schaden nehmen könnte.

[0032] An dieser Stelle sei ferner darauf hingewiesen, dass insbesondere Bedarf nach automatischen Musikanalyseverfahren besteht, ohne dass das Ergebnis immer überprüft und gegebenenfalls korrigiert werden kann. Statt dessen ist ein Verfahren nur dann am Markt einsetzbar, wenn es automatisch ohne menschliche Nachkorrektur ablaufen kann.

[0033] Die EP 1577877 A1 offenbart ein Verfahren zum automatischen Erfassen eines Refrainabschnitts, wobei hierfür mehrere Probleme gelöst werden. Das erste Problem ist das Studium akustischer Merkmale und einer Ähnlichkeit eines Abschnitts des Audiosignals mit anderen Abschnitten. Das zweite Problem ist das Kriterium, wie hoch eine Ähnlichkeit sein muss, damit ein Abschnitt als Wiederholung aufgefasst wird. Dieses Kriterium hängt von dem Audiostück selbst ab. Das dritte Problem besteht in dem Bestimmen beider Enden, also dem Beginn und dem Ende wiederholter Abschnitte, und das vierte Problem besteht in der Erfassung einer modulierten Wiederholung. Hierzu wird zunächst ein Zeit-Verzögerungs-Diagramm erstellt, indem die Ähnlichkeit zwischen einem 12-dimensionalen Chroma-Vektor eines Abschnitts mit einem entsprechenden Vektor jedes vorhergehenden Abschnitts erzeugt wird. Hierdurch entsteht für jede inkrementelle Verzögerung ein Parameter $R_{all}(t, l)$, wobei hierzu sämtliche Ähnlichkeitswerte einer Verzögerungsstufe aufsummiert werden. Dann wird zur Erfassung von Liniensegmentkandidatenspitzen ein Schwellwert bestimmt, wobei nur Spitzen in $R_{all}(t, l)$ die über dem Schwellwert liegen, ausgewählt werden. Der Schwellwert ist so eingestellt, dass eine Zwischen-Klassenverteilung bestimmt wird, wobei der Schwellwert sowohl von der Anzahl der Spitzen in jeder Klasse, der Gesamtanzahl von Spitzen und den Mittelwerten der Spitzenhöhen in jeder Klasse abhängt. Unter Verwendung der ausgewählten Kandidatenspitzen, also der Spitzen, die über dem Schwellwert liegen, werden dann integrierte Wiederholungsabschnitte bestimmt, wobei berücksichtigt wird, ob entsprechende Ähnlichkeitsliniensegmente an vorherigen Verzögerungspositionen bezüglich der Verzögerungsposition des Wiederholungsabschnitts existieren.

[0034] Die Aufgabe der vorliegenden Erfindung besteht darin, ein verbessertes und gleichzeitig effizientes Konzept zum Gruppieren von zeitlichen Segmenten eines Musikstücks zu schaffen.

[0035] Diese Aufgabe wird durch eine Vorrichtung zum Gruppieren gemäß Patentanspruch 1, ein Verfahren zum Gruppieren gemäß Patentanspruch 20 oder ein Computer-Programm gemäß Patentanspruch 21 gelöst.

[0036] Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, dass die Zuweisung eines Segments zu einer Segmentklasse auf der Basis eines adaptiven Ähnlichkeits-Mittelwerts für ein Segment durchzuführen ist, derart, dass durch den Ähnlichkeits-Mittelwert berücksichtigt wird, welchen insgesamt Ähnlichkeits-Score ein Segment im gesamten Stück hat. Nachdem für ein Segment ein solcher Ähnlichkeits-Mittelwert berechnet worden ist, zu dessen Berechnung die Anzahl der Segmente und die Ähnlichkeitswerte der dem Segment zugeordneten Mehrzahl von Ähnlichkeitswerten benötigt werden, wird dann die tatsächliche Zuweisung eines Segments zu einer Segmentklasse, also zu einem Cluster, auf der Basis dieses Ähnlichkeits-Mittelwerts durchgeführt. Liegt ein Ähnlichkeitswert eines Segments zu dem gerade betrachteten Segment beispielsweise oberhalb des Ähnlichkeits-Mittelwerts, so wird das Segment als zu der gerade betrachteten Segmentklasse zugehörig zugewiesen. Liegt der Ähnlichkeitswert eines Segments zu dem gerade betrachteten Segment dagegen unterhalb dieses Ähnlichkeits-Mittelwerts, so wird es nicht der Segmentklasse zugewiesen.

[0037] In anderen Worten ausgedrückt bedeutet dies, dass die Zuweisung nicht mehr abhängig von der absoluten Größe der Ähnlichkeitswerte durchgeführt wird, sondern relativ zu dem Ähnlichkeits-Mittelwert. Dies bedeutet, dass für ein Segment, das einen relativ geringen Ähnlichkeits-Score hat, also z. B. für ein Segment, das ein Intro oder Outro hat, der Ähnlichkeits-Mittelwert niedriger sein wird als für ein Segment, das eine Strophe oder ein Refrain ist. Damit werden die starken Abweichungen der Ähnlichkeiten von Segmenten in Stücken bzw. der Häufigkeit des Auftretens von bestimmten Segmenten in Stücken berücksichtigt, wobei z. B. numerische Probleme und damit auch Mehrdeutigkeiten und damit verbundene Falsch-Zuweisungen vermieden werden können.

[0038] Das erfindungsgemäße Konzept eignet sich besonders für Musikstücke, die nicht nur aus Strophen und Refrains bestehen, also die Segmente haben, die zu Segmentklasse gehören, die gleich große Ähnlichkeitswerte haben, sondern auch für Stücke, die neben Strophe und Refrain auch andere Teile haben, nämlich eine Einführung (Intro), ein Zwischenspiel (Bridge) oder einen Ausklang (Outro).

[0039] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung werden die Berechnung des adaptiven Ähnlichkeits-Mittelwertes und das Zuweisen eines Segments iterativ durchgeführt, wobei zugewiesene Segmente beim nächsten Iterationsdurchgang ignoriert werden. Damit ändert sich für den nächsten Iterationsdurchgang wieder der Ähnlichkeits-Absolutwert, also die Summe der Ähnlichkeitswerte in einer Spalte der Ähnlichkeitsmatrix, da bereits zugewiesene Segmente zu 0 gesetzt worden sind.

[0040] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird eine Segmentierungs-Nachkorrektur durchgeführt, und zwar dahingehend, dass nach der Segmentierung z. B. auf Grund des Neuheitswertes (der lokalen Maxima des Neuheitswertes) und nach einer anschließenden Zuordnung zu Segmentklassen relativ kurze Segmente untersucht werden, um zu sehen, ob sie dem Vorgänger-Segment oder dem Nachfolger-Segment zugeordnet werden können, da Segmente unterhalb einer minimalen Segmentlänge mit hoher Wahrscheinlichkeit auf eine Übersegmentierung hindeuten.

[0041] Bei einem alternativen bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird nach der abschließenden Segmentierung und Zuordnung in die Segmentklassen ein Labeling durchgeführt, und zwar unter Verwendung eines speziellen Auswahlalgorithmus, um eine möglichst korrekte Kennzeichnung der Segmentklassen als Strophe oder Refrain zu erhalten.

[0042] Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend Bezug nehmend auf die beiliegende Zeichnung detailliert erläutert. Es zeigen:

Fig. 1 ein Blockschaltbild der erfindungsgemäßen Vorrichtung zum Gruppieren gemäß einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung;

Fig. 2 ein Flussdiagramm zur Darstellung einer bevorzugten Ausführungsform der Erfindung zum iterativen Zuweisen;

Fig. 3 ein Blockdiagramm der Funktionsweise der Segmentierungskorrektureinrichtung;

Fig. 4a und Fig. 4b eine bevorzugte Ausführungsform der Segmentklassenbezeichnungseinrichtung;

Fig. 5 ein Gesamtschaltbild eines Audioanalysewerkzeugs;

Fig. 6 eine Darstellung einer beispielhaften Merkmalsähnlichkeitsmatrix;

Fig. 7 eine beispielhafte Darstellung einer Segmentähnlichkeitsmatrix;

Fig. 8 eine schematische Darstellung zur Veranschaulichung der Elemente in einer Ähnlichkeitsmatrix

S; und

Fig. 9 eine schematische Darstellung eines geglätteten Neuheitswerts.

- 5 **[0043]** Fig. 1 zeigt eine Vorrichtung zum Gruppieren von zeitlichen Segmenten eines Musikstücks, das in in dem Musikstück wiederholt auftretende Hauptteile gegliedert ist, in verschiedene Segmentklassen, wobei eine Segmentklasse einem Hauptteil zugeordnet ist. Die vorliegende Erfindung bezieht sich somit besonders auf Musikstücke, die einer gewissen Struktur unterliegen, in der ähnliche Abschnitte mehrmals erscheinen und sich mit anderen Abschnitten abwechseln. Die meisten Rock- und Popsongs besitzen eine klare Struktur in Bezug auf ihre Hauptteile.
- 10 **[0044]** Die Literatur behandelt das Thema der Musikanalyse hauptsächlich anhand klassischer Musik, davon gilt jedoch auch vieles für Rock- und Popmusik. Die Hauptteile eines Musikstücks werden auch "Großformteile" genannt. Unter einem Großformteil eines Stücks versteht man einen Abschnitt, der hinsichtlich verschiedener Merkmale, z. B. Melodik, Rhythmik, Textur, usw., eine relativ einheitliche Beschaffenheit hat. Diese Definition gilt allgemein in der Musiktheorie.
- 15 **[0045]** Großformteile in der Rock- und Popmusik sind z. B. Strophe, Refrain, Bridge und Solo. In der klassischen Musik wird ein Wechselspiel von Refrain und anderen Teilen (Couplets) einer Komposition auch Rondo genannt. Im Allgemeinen kontrastieren die Couplets zum Refrain, beispielsweise hinsichtlich Melodik, Rhythmik, Harmonik, Tonart oder Instrumentation. Dies lässt sich auch auf moderne Unterhaltungsmusik übertragen. So wie es bei dem Rondo verschiedene Formen gibt (Kettenrondo, Bogenrondo, Sonatenrondo), bestehen auch in Rock- und Popmusik bewährte Muster zum Aufbau eines Liedes. Diese sind natürlich nur einige Möglichkeiten aus vielen. Letztendlich entscheidet natürlich der
- 20 Komponist, wie sein Stück aufgebaut ist. Ein Beispiel für einen typischen Aufbau eines Rockliedes ist das Muster.

A-B-A-B-C-D-A-B,

- 25 wobei A gleich Strophe, B gleich Refrain, C gleich Bridge und D gleich Solo gilt. Oftmals wird ein Musikstück mit einem Vorspiel (Intro) eingeleitet. Intros bestehen häufig aus der gleichen Akkordfolge wie die Strophe, allerdings mit anderer Instrumentation, z. B. ohne Schlagzeug, ohne Bass oder ohne Verzerrung der Gitarre bei Rockliedern etc.

- [0046]** Die erfindungsgemäße Vorrichtung umfasst zunächst eine Einrichtung 10 zum Bereitstellen einer Ähnlichkeitsdarstellung für die Segmente, wobei die Ähnlichkeitsdarstellung für jedes Segment eine zugeordnete Mehrzahl von Ähnlichkeitswerten aufweist, wobei die Ähnlichkeitswerte angeben, wie ähnlich das Segment zu jedem anderen Segment ist. Die Ähnlichkeitsdarstellung ist vorzugsweise die in Fig. 7 gezeigte Segment-Ähnlichkeitsmatrix. Sie hat für jedes Segment (in Fig. 7 Segmente 1-10) eine eigene Spalte, die den Index "j" hat. Ferner hat die Ähnlichkeitsdarstellung für jedes Segment eine eigene Zeile, wobei eine Zeile mit einem Zeilenindex i bezeichnet ist. Dies wird nachfolgend anhand des beispielhaften Segments 5 bezeichnet. Das Element (5,5) in der Hauptdiagonale der Matrix von Fig. 7 ist der Ähnlichkeitswert des Segments 5 mit sich selbst, also der maximale Ähnlichkeitswert. Ferner ist das Segment 5 noch mittelähnlich zum Segment Nr. 6, wie es durch das Element (6,5) oder durch das Element (5,6) der Matrix in Fig. 7 bezeichnet ist. Darüber hinaus hat das Segment 5 noch Ähnlichkeiten zu den Segmenten 2 und 3, wie es durch die Elemente (2,5) oder (3,5) oder (5,2) oder (5,3) in Fig. 7 gezeigt ist. Zu den anderen Segmenten 1, 4, 7, 8, 9, 10 hat das Segment Nr. 5 eine Ähnlichkeit, die in Fig. 7 nicht mehr sichtbar ist.

- [0047]** Eine dem Segment zugeordnete Mehrzahl von Ähnlichkeitswerten ist beispielsweise eine Spalte oder eine Zeile der Segment-Ähnlichkeitsmatrix in Fig. 7, wobei diese Spalte oder Zeile auf Grund ihres Spalten-/Zeilen-Indexes angibt, auf welches Segment sie sich bezieht, nämlich beispielsweise auf das fünfte Segment, und wobei diese Zeile/Spalte die Ähnlichkeiten des fünften Segments zu jedem anderen Segment in dem Stück umfasst. Die Mehrzahl von Ähnlichkeitswerten ist also beispielsweise eine Zeile der Ähnlichkeitsmatrix oder, alternativ, eine Spalte der Ähnlichkeitsmatrix von Fig. 7.

- 45 **[0048]** Die Vorrichtung zum Gruppieren von zeitlichen Segmenten des Musikstücks umfasst ferner eine Einrichtung 12 zum Berechnen eines Ähnlichkeits-Mittelwertes für ein Segment, unter Verwendung der Segmente und der Ähnlichkeitswerte der dem Segment zugeordneten Mehrzahl von Ähnlichkeitswerten. Die Einrichtung 12 ist ausgebildet, um z. B. für die Spalte 5 in Fig. 7 einen Ähnlichkeits-Mittelwert zu berechnen. Wird bei einem bevorzugten Ausführungsbeispiel der arithmetische Mittelwert verwendet, so wird die Einrichtung 12 die Ähnlichkeitswerte in der Spalte addieren und durch die Anzahl der Segmente insgesamt teilen. Um die Selbstähnlichkeit zu eliminieren, könnte von dem Additionsergebnis auch die Ähnlichkeit des Segments mit sich selbst abgezogen werden, wobei natürlich dann auch eine Teilung nicht mehr durch sämtliche Elemente, sondern durch sämtliche Elemente weniger 1 durchzuführen ist.

- [0049]** Die Einrichtung 12 zum Berechnen könnte alternativ auch den geometrischen Mittelwert berechnen, also jeden Ähnlichkeitswert einer Spalte für sich quadrieren, um die quadrierten Ergebnisse zu summieren, um dann eine Wurzel aus dem Summationsergebnis zu berechnen, welche durch die Anzahl der Elemente in der Spalte (bzw. die Anzahl der Elemente in der Spalte weniger 1) zu teilen ist. Beliebige andere Mittelwerte, wie beispielsweise der Median-Wert etc. sind verwendbar, so lange der Mittelwert für jede Spalte der Ähnlichkeitsmatrix adaptiv berechnet wird, also ein Wert ist, der unter Verwendung der Ähnlichkeitswerte der dem Segment zugeordneten Mehrzahl von Ähnlichkeitswerten

berechnet wird.

[0050] Der adaptiv berechnete Ähnlichkeits-Schwellenwert wird dann einer Einrichtung 14 zum Zuweisen eines Segments zu einer Segmentklasse geliefert. Die Einrichtung 14 zum Zuweisen ist ausgebildet, um ein Segment einer Segmentklasse zuzuordnen, wenn der Ähnlichkeitswert des Segments eine vorbestimmte Bedingung bezüglich des Ähnlichkeits-Mittelwerts erfüllt. Ist beispielsweise der Ähnlichkeits-Mittelwert so, dass ein größerer Wert auf eine größere Ähnlichkeit und ein kleinerer Wert auf eine geringere Ähnlichkeit hinweist, so wird die vorbestimmte Beziehung darin bestehen, dass der Ähnlichkeitswert eines Segments gleich oder oberhalb des Ähnlichkeits-Mittelwerts sein muss, damit das Segment einer Segmentklasse zugewiesen wird.

[0051] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung existieren noch weitere Einrichtungen, um spezielle Ausführungsformen zu verwirklichen, auf die später eingegangen wird. Diese Einrichtungen sind eine Segmentauswahleinrichtung 16, eine Segmentzuweisungskonflikteinrichtung 18, eine Segmentierungskorrektureinrichtung 20 sowie eine Segmentklassen-Bezeichnungseinrichtung 22.

[0052] Die Segmentauswahleinrichtung 16 in Fig. 1 ist ausgebildet, um zunächst für jede Spalte in der Matrix von Fig. 7 einen Gesamt-Ähnlichkeitswert $V(j)$ zu berechnen, der folgendermaßen ermittelt wird:

$$V(j) = \sum_{i=1}^P S_s(i, j) - S\ddot{A} \quad j = 1, \dots, P$$

P ist die Anzahl der Segmente. $S\ddot{A}$ ist der Wert der Selbstähnlichkeit eines Segments mit sich selbst. Je nach verwendeter Technik kann der Wert z. B. Null oder Eins sein. Die Segmentauswahleinrichtung 16 wird zunächst den Wert $V(j)$ für jedes Segment berechnen, um dann das Vektorelement i des Vektors V mit maximalem Wert herauszufinden. Anders ausgedrückt bedeutet dies, dass die Spalte in Fig. 7 gewählt wird, die bei der Aufaddition der einzelnen Ähnlichkeitswerte in der Spalte den größten Wert oder Score erreicht. Dieses Segment könnte beispielsweise das Segment Nr. 5 bzw. die Spalte 5 der Matrix in Fig. 7 sein, da dieses Segment mit drei anderen Segmenten zumindest eine gewisse Ähnlichkeit hat. Ein anderer Kandidat bei dem Beispiel von Fig. 7 könnte auch das Segment mit der Nr. 7 sein, da dieses Segment ebenfalls zu drei anderen Segmenten eine gewisse Ähnlichkeit hat, die zudem noch größer ist als die Ähnlichkeit des Segments 5 zu den Segmenten 2 und 3 (höhere Grauschattierung in Fig. 7).

[0053] Für das nachfolgende Beispiel wird nunmehr angenommen, dass die Segmentauswahleinrichtung 16 das Segment Nr. 7 auswählt, da es den höchsten Ähnlichkeits-Score auf Grund der Matrixelemente (1,7), (4,7) und (10,7) hat. Dies bedeutet in anderen Worten, dass $V(7)$ die Komponente des Vektors V ist, die den maximalen Wert unter allen Komponenten von V hat.

[0054] Nunmehr wird der Ähnlichkeits-Score der Spalte 7, also für das Segment Nr. 7 noch durch die Zahl "9" geteilt, um aus der Einrichtung 12 den Ähnlichkeits-Schwellwert für das Segment zu erhalten.

[0055] Hierauf wird in der Segment-Ähnlichkeitsmatrix für die siebte Zeile bzw. Spalte überprüft, welche Segmentähnlichkeiten über dem berechneten Schwellwert liegen, d. h. mit welchen Segmenten das i -te Segment eine überdurchschnittliche Ähnlichkeit aufweist. All diese Segmente werden nunmehr ebenfalls wie das siebte Segment einer ersten Segmentklasse zugewiesen.

[0056] Für das vorliegende Beispiel wird angenommen, dass die Ähnlichkeit des Segments 10 zum Segment 7 unterdurchschnittlich ist, dass jedoch die Ähnlichkeiten des Segments 4 und des Segments 1 zum Segment 7 überdurchschnittlich sind. Daher werden in die erste Segmentklasse neben dem Segment Nr. 7 auch das Segment Nr. 4 und das Segment Nr. 1 eingeordnet. Dagegen wird das Segment Nr. 10 auf Grund der unterdurchschnittlichen Ähnlichkeit zum Segment Nr. 7 nicht in die erste Segmentklasse eingeordnet.

[0057] Nach der Zuweisung werden die korrespondierenden Vektorelemente $V(j)$ aller Segmente, die in dieser Schwellwertuntersuchung einem Cluster zugeordnet wurden, zu 0 gesetzt. Im Beispiel sind dies neben $V(7)$ auch die Komponenten $V(4)$ und $V(1)$. Dies bedeutet unmittelbar, dass die 7., 4. und 1. Spalte der Matrix nicht mehr für eine spätere Maximalsuche zur Verfügung stehen werden, da sie Null sind, also keinesfalls ein Maximum sein können.

[0058] Dies ist in etwa gleichbedeutend damit, dass die Einträge (1,7), (4,7), (7,7) und (10,7) der Segmentähnlichkeitsmatrix zu 0 gesetzt werden. Dasselbe Prozedere wird für die Spalte 1 (Elemente (1,1), (4,1) und (7,1)) und die Spalte 4 (Elemente (1,4), (4,4), (7,4) und (10,4)) durchgeführt. Aufgrund der einfacheren Handhabbarkeit, wird jedoch die Matrix nicht verändert, sondern werden die Komponenten von V , die zu einem zugewiesenen Segment gehören, bei der nächsten Maximalsuche in einem späteren Iterationsschritt ignoriert.

[0059] In einem nächsten Iterationsschritt wird nunmehr ein neues Maximum unter den noch verbleibenden Elementen von V , also unter $V(2)$, $V(3)$, $V(5)$, $V(6)$, $V(8)$, $V(9)$ und $V(10)$ gesucht. Voraussichtlich wird dann das Segment Nr. 5, also $V(5)$, den größten Ähnlichkeits-Score ergeben. Die zweite Segmentklasse erhält dann die Segmente 5 und 6. Auf Grund der Tatsache, dass die Ähnlichkeiten zu den Segmenten 2 und 3 unterdurchschnittlich sind, werden die Segmente 2 und 3 nicht in den Cluster zweiter Ordnung gebracht. Damit werden die Elemente $V(6)$ und $V(5)$ vom Vektor V auf

Grund der erfolgten Zuweisung zu 0 gesetzt, während noch die Komponenten V(2), V(3), V(8), V(9) und V(10) des Vektors für die Auswahl des Clusters dritter Ordnung verbleiben.

[0060] Hierauf wird wieder ein neues Maximum unter den genannten verbleibenden Elementen von V gesucht. Das neue Maximum könnte V(10) sein, also die Komponente von V für das Segment 10. Segment 10 kommt also in die Segmentklasse dritter Ordnung. So könnte sich ferner herausstellen, dass das Segment 7 auch zum Segment 10 eine überdurchschnittliche Ähnlichkeit hat, obgleich das Segment 7 bereits der ersten Segmentklasse zugehörig gekennzeichnet ist. Es entsteht somit ein Zuweisungskonflikt, der durch die Segmentzuweisungskonflikteinrichtung 18 von Fig. 1 aufgelöst wird.

[0061] Eine einfache Art der Auflösung könnte sein, einfach in die dritte Segmentklasse das Segment 7 nicht zuzuweisen und z. B. statt dessen das Segment 4 zuzuweisen, falls für das Segment 4 nicht ebenfalls ein Konflikt existieren würde.

[0062] Vorzugsweise wird jedoch, um die Ähnlichkeit zwischen dem Segment 7 und dem Segment 10 nicht unberücksichtigt zu lassen, die Ähnlichkeit zwischen 7 und 10 in nachfolgendem Algorithmus berücksichtigt.

[0063] Allgemein ist die Erfindung ausgelegt, um die Ähnlichkeit zwischen i und k nicht unberücksichtigt zu lassen. Daher werden die Ähnlichkeitswerte $S_s(i,k)$ von Segment i und k mit dem Ähnlichkeitswert $S_s(i^*,k)$ verglichen, wobei i^* das erste Segment ist, das dem Cluster C^* zugeordnet wurde. Der Cluster bzw. die Segmentklasse C^* ist der Cluster, dem das Segment k bereits auf Grund einer vorherigen Untersuchung zugeordnet ist. Der Ähnlichkeitswert $S_s(i^*,k)$ ist ausschlaggebend dafür, dass das Segment k dem Cluster C^* zugehörig ist. Ist $S_s(i^*,k)$ größer als $S_s(i,k)$, so bleibt das Segment k im Cluster C^* . Ist $S_s(i^*,k)$ kleiner als $S_s(i,k)$, so wird das Segment k aus dem Cluster C^* herausgenommen und dem Cluster C zugewiesen. Für den ersten Fall, also wenn das Segment k nicht die Clusterzugehörigkeit wechselt, wird für das Segment i eine Tendenz zum Cluster C^* vermerkt. Vorzugsweise wird diese Tendenz jedoch auch dann vermerkt, wenn das Segment k die Clusterzugehörigkeit wechselt. In diesem Fall wird eine Tendenz dieses Segments zum Cluster, in den es ursprünglich aufgenommen wurde, vermerkt. Diese Tendenzen können vorteilhafterweise bei einer Segmentierungskorrektur, die durch die Segmentierungskorrektureinrichtung 20 ausgeführt wird, verwendet werden.

[0064] Die Ähnlichkeitswertüberprüfung wird, auf Grund der Tatsache, dass das Segment 7 das "Ursprungssegment" in der ersten Segmentklasse ist, zugunsten der ersten Segmentklasse ausgehen. Das Segment 7 wird also seine Clusterzugehörigkeit (Segmentzugehörigkeit) nicht ändern, sondern es wird in der ersten Segmentklasse verbleiben. Diese Tatsache wird jedoch dadurch berücksichtigt, dass dem Segment Nr. 10 in der dritten Segmentklasse ein Trend zur ersten Segmentklasse attestiert wird.

[0065] Erfindungsgemäß wird damit berücksichtigt, dass insbesondere für die Segmente, deren Segment-Ähnlichkeiten zu zwei unterschiedlichen Segmentklassen existieren, diese Ähnlichkeiten dennoch nicht ignoriert werden, sondern gegebenenfalls später durch den Trend bzw. die Tendenz doch noch berücksichtigt werden.

[0066] Das Prozedere wird so lange fortgeführt, bis alle Segmente in der Segment-Ähnlichkeitsmatrix zugeordnet sind, was der Fall ist, wenn alle Elemente vom Vektor V zu Null gesetzt sind.

[0067] Dies würde für das in Fig. 7 gezeigte Beispiel bedeuten, dass als nächstes, in die vierte Segmentklasse, das Maximum von V(2), V(3), V(8), V(9), also das Segment 2 und 3 eingeordnet werden, um dann, in einer fünften Segmentklasse, die Segmente 8 bzw. 9 einzuordnen, bis alle Segmente zugeordnet worden sind. Damit ist der in Fig. 2 gezeigte iterative Algorithmus beendet.

[0068] Nachfolgend wird detailliert auf die bevorzugte Implementierung der Segmentierungskorrektureinrichtung 20 anhand von Fig. 3 eingegangen.

[0069] So ergibt sich, dass bei der Berechnung der Segmentgrenzen mittels der Kernel-Korrelation, jedoch auch bei der Berechnung von Segmentgrenzen mittels anderer Maßnahmen häufig eine Übersegmentierung eines Stücks entsteht, d. h. es werden zu viele Segmentgrenzen bzw. allgemein zu kurze Segmente berechnet. Eine Übersegmentierung, z. B. hervorgerufen durch eine falsche Unterteilung der Strophe, wird erfindungsgemäß dadurch korrigiert, dass auf Grund der Segmentlänge und der Information, in welche Segmentklasse ein Vorgänger- oder Nachfolger-Segment einsortiert worden ist, korrigiert wird. In anderen Worten ausgedrückt dient die Korrektur dazu, zu kurze Segmente vollständig zu eliminieren, also mit benachbarten Segmenten zu verschmelzen, und um Segmente, die kurz sind, jedoch nicht zu kurz sind, also die eine kurze Länge haben, jedoch länger als die Minimallänge sind, noch einer besonderen Untersuchung zu unterziehen, ob sie vielleicht nicht doch noch mit einem Vorgänger-segment oder einem Nachfolger-segment verschmolzen werden können. Grundsätzlich werden erfindungsgemäß aufeinander folgende Segmente, die der gleichen Segmentklasse angehören, immer verschmolzen. Ergibt das in Fig. 7 gezeigte Szenario z. B., dass die Segmente 2 und 3 in dieselbe Segmentklasse kommen, so werden diese automatisch miteinander verschmolzen, während die Segmente in der ersten Segmentklasse, also die Segmente 7, 4, 1 voneinander beabstandet sind und daher (wenigstens zunächst) nicht verschmelzbar sind. Dies wird in Fig. 3 durch einen Block 30 angedeutet. Nunmehr wird in einem Block 31 untersucht, ob Segmente eine Segmentlänge haben, die kleiner als eine Mindestlänge ist. So existieren vorzugsweise verschiedene Mindestlängen.

[0070] Es werden relativ kurze Segmente, die kürzer als 11 Sekunden (eine erste Schwelle) sind, überhaupt nur

untersucht, während später noch kürzere Segmente (eine zweite Schwelle, die kleiner als die erste ist), die kürzer als 9 Sekunden sind, untersucht werden, und später noch verbleibende Segmente, die kürzer als 6 Sekunden (eine dritte Schwelle, die kürzer als die zweite Schwelle ist) sind, wieder alternativ behandelt werden.

[0071] Bei dem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung, bei dem diese gestaffelte Längenüberprüfung stattfindet, ist die Segmentlängenüberprüfung im Block 31 zunächst darauf gerichtet, dass die Segmente kürzer als 11 Sekunden gefunden werden. Für die Segmente, die länger als 11 Sekunden sind, wird keine Nachbearbeitung vorgenommen, wie es durch ein "Nein" am Block 31 erkennbar ist. Für Segmente, die kürzer als 11 Sekunden sind, wird zunächst eine Tendenzenüberprüfung (Block 32) durchgeführt. So wird zunächst untersucht, ob ein Segment auf Grund der Funktionalität der Segmentzuweisungskonflikteinrichtung 18 von Fig. 1 einen zugeordneten Trend bzw. eine zugeordnete Tendenz hat. Bei dem Beispiel von Fig. 7 wäre dies das Segment 10, das einen Trend zu dem Segment 7 bzw. einen Trend zur ersten Segmentklasse hat. Ist das zehnte Segment kürzer als 11 Sekunden, so würde bei dem in Fig. 7 gezeigten Beispiel dennoch auch auf Grund der Tendenzüberprüfung nichts geschehen, da eine Verschmelzung des betrachteten Segments nur dann stattfindet, wenn es eine Tendenz nicht zu irgendeinem Cluster, also zu irgendeiner Segmentklasse hat, sondern eine Tendenz zu einem Cluster eines angrenzenden (vorher oder nachher) Segments. Dies ist jedoch für das Segment 10 bei dem in Fig. 7 gezeigten Beispiel nicht der Fall.

[0072] Um auch die zu kurzen Segmente zu vermeiden, die keine Tendenz zu dem Cluster eines benachbarten Segments aufweisen, wird vorgegangen, wie es in den Blöcken 33a, 33b, 33c und 33d in Fig. 3 dargestellt ist. So wird an Segmenten, die länger als 9 Sekunden, jedoch kürzer als 11 Sekunden sind, nichts mehr gemacht. Sie werden belassen. In einem Block 33a wird nun jedoch für Segmente aus dem Cluster X, die kürzer als 9 Sekunden sind, und bei denen sowohl das Vorgänger-Segment als auch das Nachfolge-Segment zum Cluster Y gehören, eine Zuweisung zum Cluster Y vorgenommen, was automatisch bedeutet, dass ein solches Segment mit sowohl dem Vorgänger- als auch dem Nachfolger-Segment verschmolzen wird, so dass ein insgesamt längeres Segment entsteht, das sich aus dem betrachteten Segment sowie dem Vorgänger- als auch dem Nachfolger-Segment zusammensetzt. Somit kann durch eine nachfolgende Verschmelzung eine Zusammenfassung von zunächst getrennten Segmenten über ein zu verschmelzendes dazwischenliegendes Segment gelingen.

[0073] In einem Block 33b ist ferner ausgeführt, was mit einem Segment geschieht, das kürzer als 9 Sekunden ist, und das das einzige Segment in einer Segmentgruppe ist. So ist in der dritten Segmentklasse das Segment Nr. 10 das einzige Segment. Wäre es nun kürzer als 9 Sekunden, so wird es automatisch der Segmentklasse zugeordnet, zu der das Segment Nr. 9 gehört. Dies führt automatisch zu einer Verschmelzung des Segments 10 mit dem Segment 9. Ist das Segment 10 länger als 9 Sekunden, so wird diese Verschmelzung nicht vorgenommen.

[0074] In einem Block 33c wird dann eine Untersuchung vorgenommen für Segmente, die kürzer als 9 Sekunden sind, und die nicht das einzige Segment in einem entsprechenden Cluster X, als in einer entsprechenden Segmentgruppe sind. Sie werden einer genaueren Überprüfung unterzogen, in der eine Regelmäßigkeit in der Clusterabfolge festgestellt werden soll. Zunächst werden alle Segmente aus der Segmentgruppe X gesucht, die kürzer als die Mindestlänge sind. Im Anschluss wird für jedes dieser Segmente geprüft, ob die Vorgänger- und Nachfolge-Segmente jeweils zu einem einheitlichen Cluster gehören. Sind alle Vorgänger-Segmente aus einem einheitlichen Cluster, so werden alle zu kurzen Segmente aus dem Cluster X dem Vorgänger-Cluster zugeordnet. Sind dagegen alle Nachfolger-Segmente aus einem einheitlichen Cluster, werden die zu kurzen Segmente aus dem Cluster X jeweils dem Nachfolger-Cluster zugeordnet.

[0075] In einem Block 33d ist ausgeführt, was passiert, wenn auch diese Bedingung für Segmente nicht erfüllt ist, die kürzer als 9 Sekunden sind. In diesem Fall wird eine Neuheitswertüberprüfung durchgeführt, indem auf die Neuheitswertkurve zurückgegriffen wird, die in Fig. 9 dargestellt ist. Insbesondere wird die Neuheitskurve, die aus der Kernel-Korrelation entstanden ist, an den Stellen der betroffenen Segmentgrenzen ausgelesen, und das Maximum dieser Werte wird ermittelt. Tritt das Maximum an einem Segmentanfang auf, werden die zu kurzen Segmente dem Cluster des Nachfolge-Segments zugeordnet. Tritt das Maximum an einem Segmentende auf, werden die zu kurzen Segmente dem Cluster des Vorgänger-Segments zugeordnet. Wäre das in Fig. 9 mit 90 bezeichnete Segment ein Segment, das kürzer als 9 Sekunden ist, so würde die Neuheitsüberprüfung am Anfang des Segments 90 einen höheren Neuheitswert 91 ergeben als am Ende des Segments, wobei der Neuheitswert am Ende des Segments mit 92 bezeichnet ist. Dies würde bedeuten, dass das Segment 90 dem Nachfolger-Segment zugeordnet werden würde, da der Neuheitswert zum Nachfolger-Segment geringer ist als der Neuheitswert zum Vorgänger-Segment.

[0076] Verbleiben nunmehr noch Segmente, die kürzer als 9 Sekunden sind und noch nicht verschmolzen werden durften, so wird unter diesen noch einmal eine gestaffelte Auswahl durchgeführt. Insbesondere werden nunmehr alle Segmente unter den verbleibenden Segmenten, die kürzer als 6 Sekunden sind, ausgewählt. Die Segmente, deren Länge zwischen 6 und 9 Sekunden aus dieser Gruppe sind, werden "unangetastet" zugelassen.

[0077] Die Segmente, die kürzer als 6 Sekunden sind, werden nunmehr jedoch alle der anhand der Elemente 90, 91, 92 erklärten Neuheitsprüfung unterzogen und entweder dem Vorgänger- oder dem Nachfolger-Segment zugeordnet, so dass am Ende des in Fig. 3 gezeigten Nachkorrekturalgorithmus alle zu kurzen Segmente, nämlich alle Segmente unterhalb einer Länge von 6 Sekunden, mit Vorgänger- und Nachfolger-Segmenten intelligent verschmolzen worden sind.

[0078] Dieses erfindungsgemäße Prozedere hat den Vorteil, dass keine Elimination von Teilen des Stücks durchgeführt

worden ist, dass also keine einfache Eliminierung der zu kurzen Segmente durch Zu-Null-Setzen durchgeführt worden ist, sondern dass nach wie vor das gesamte komplette Musikstück durch die Gesamtheit der Segmente repräsentiert ist. Durch die Segmentierung ist daher kein Informationsverlust aufgetreten, der jedoch sein würde, wenn man z. B. als Reaktion auf die Übersegmentierung einfach alle zu kurzen Segmente "ohne Rücksicht auf Verluste" einfach eliminieren würde.

[0079] Nachfolgend wird Bezug nehmend auf Fig. 4a und Fig. 4b eine bevorzugte Implementierung der Segmentklassen-Bezeichnungseinrichtung 22 von Fig. 1 dargestellt. Erfindungsgemäß werden beim Labeling zwei Clustern die Label "Strophe" und "Refrain" zugewiesen.

[0080] Erfindungsgemäß wird nun nicht etwa ein größter Singularwert einer Singularwertzerlegung und der dazugehörige Cluster als Refrain und der Cluster für den zweitgrößten Singularwert als Strophe verwendet. Ferner wird nicht grundsätzlich davon ausgegangen, dass jedes Lied mit einer Strophe anfängt, dass also der Cluster mit dem ersten Segment der Strophenclass ist und der andere Cluster der Refraincluster ist. Statt dessen wird erfindungsgemäß der Cluster in der Kandidatenauswahl, der das letzte Segment hat, als Refrain bezeichnet, und der andere Cluster wird als Strophe bezeichnet.

[0081] So wird also für die beiden letztendlich zur Strophe-/Refrain-Auswahl bereitstehenden Cluster überprüft (40), welcher Cluster das Segment hat, das als letztes Segment der Segmente der beiden Segmentgruppen im Liedverlauf vorkommt, um denselben als Refrain zu bezeichnen.

[0082] Das letzte Segment kann das tatsächlich im Lied letzte Segment sein oder aber ein Segment, das im Lied später auftritt als alle Segmente der anderen Segmentklasse. Ist dieses Segment nicht das tatsächlich letzte Segment im Lied, so bedeutet dies, dass noch ein Outro vorliegt.

[0083] Diese Entscheidung basiert auf der Erkenntnis, dass der Refrain in den allermeisten Fällen in einem Lied hinter der letzten Strophe kommt, also direkt als letztes Segment des Liedes, wenn ein Stück z. B. mit dem Refrain ausgeblendet wird, oder als Segment vor einem Outro, das auf einen Refrain folgt und mit dem das Stück beendet wird.

[0084] Ist das letzte Segment aus der ersten Segmentgruppe, dann werden alle Segmente dieser ersten (höchstwertigen) Segmentklasse als Refrain bezeichnet, wie es durch einen Block 41 in Fig. 4b dargestellt ist. Zusätzlich werden in diesem Fall alle Segmente der anderen Segmentklasse, die zur Auswahl steht, als "Strophe" gekennzeichnet, da typischerweise von den beiden Kandidaten-Segmentklassen eine Klasse der Refrain und damit unmittelbar die andere Klasse die Strophen haben wird.

[0085] Ergibt dagegen die Untersuchung im Block 40, nämlich welche Segmentklasse in der Auswahl das letzte Segment im Musikstückverlauf hat, dass dies die zweite, also eher niederwertige Segmentklasse ist, so wird in einem Block 42 untersucht, ob die zweite Segmentklasse das erste Segment im Musikstück hat. Diese Untersuchung basiert auf der Erkenntnis, dass die Wahrscheinlichkeit sehr hoch ist, dass ein Lied mit einer Strophe, und nicht mit einem Refrain anfängt.

[0086] Wird die Frage im Block 42 mit "Nein" beantwortet, hat also die zweite Segmentklasse nicht das erste Segment im Musikstück, so wird die zweite Segmentklasse als Refrain bezeichnet, und wird die erste Segmentklasse als Strophe bezeichnet, wie es in einem Block 43 angedeutet ist. Wird dagegen die Abfrage im Block 42 mit "Ja" beantwortet, so wird entgegen der Regel die zweite Segmentgruppe als Strophe und die erste Segmentgruppe als Refrain bezeichnet, wie es in einem Block 44 angedeutet ist. Die Bezeichnung im Block 44 geschieht deswegen, da die Wahrscheinlichkeit, dass die zweite Segmentklasse dem Refrain entspricht, schon recht gering ist. Kommt nun noch die Unwahrscheinlichkeit hinzu, dass ein Musikstück mit einem Refrain eingeleitet wird, so deutet einiges auf einen Fehler im Clustering hin, z. B. dass das zuletzt betrachtete Segment fälschlicherweise der zweiten Segmentklasse zugeordnet wurde.

[0087] In Fig. 4b wurde dargestellt, wie anhand von zwei zur Verfügung stehenden Segmentklassen die Strophe-/Refrain-Bestimmung durchgeführt worden ist. Nach dieser Strophe-/Refrain-Bestimmung können dann die restlichen Segmentklassen in einem Block 45 bezeichnet werden, wobei ein Outro gegebenenfalls die Segmentklasse sein wird, die das letzte Segment des Stücks an sich hat, während ein Intro die Segmentklasse sein wird, die das erste Segment eines Stücks an sich hat.

[0088] Nachfolgend wird anhand von Fig. 4a dargestellt, wie die beiden Segmentklassen ermittelt werden, die die Kandidaten für den in Fig. 4b gezeigten Algorithmus abgeben.

[0089] Allgemein wird im Labeling eine Zuweisung der Label "Strophe" und "Refrain" durchgeführt, wobei eine Segmentgruppe als Strophen-Segmentgruppe markiert wird, während die andere Segmentgruppe als Refrain-Segmentgruppe markiert wird. Grundsätzlich basiert dieses Konzept auf der Annahme (A1), dass die beiden Cluster (Segmentgruppen) mit den höchsten Ähnlichkeitswerten, also Cluster 1 und Cluster 2, den Refrain- und Strophenclassern entsprechen. Das von diesen beiden Clustern als letztes auftretende ist das Refrain-Cluster, wobei davon ausgegangen wird, dass eine Strophe auf einen Refrain folgt.

[0090] Die Erfahrung aus zahlreichen Tests hat gezeigt, dass Cluster 1 in den meisten Fällen dem Refrain entspricht. Für Cluster 2 wird die Annahme (A1) jedoch oftmals nicht erfüllt. Diese Situation tritt meistens dann auf, wenn es entweder noch einen dritten, sich häufig wiederholenden Teil im Stück gibt, z. B. eine Bridge, bei einer hohen Ähnlichkeit von Intro und Outro, oder aber für den nicht selten auftretenden Fall, dass ein Segment im Stück eine hohe Ähnlichkeit zum

Refrain aufweist, somit auch eine hohe Gesamtähnlichkeit hat, die Ähnlichkeit zum Refrain aber gerade nicht groß genug ist, um noch zum Cluster 1 zu gehören.

[0091] Untersuchungen haben gezeigt, dass diese Situation häufig für Abwandlungen des Refrains am Ende des Stücks auftritt. Um mit möglichst hoher Sicherheit Refrain und Strophe richtig zu markieren (labeln), wird die in Fig. 4b beschriebene Segmentauswahl dahingehend verbessert, dass, wie es in Fig. 4a dargestellt wird, die beiden Kandidaten für die Strophe-Refrain-Auswahl abhängig von den in denselben vorhandenen Segmenten bestimmt wird.

[0092] Zunächst wird in einem Schritt 46 der Cluster bzw. die Segmentgruppe mit höchstem Ähnlichkeitswert (Wert der Komponente von V, die einmal ein Maximum für die zuerst bestimmte Segmentklasse, also Segment 7 bei dem Beispiel von Fig. 7, war), also die Segmentgruppe, die beim ersten Durchlauf von Fig. 1 ermittelt worden ist, in die Strophe-Refrain-Auswahl als erster Kandidat einbezogen.

[0093] Fraglich ist nunmehr, welche weitere Segmentgruppe der zweite Teilnehmer an der Strophe-Refrain-Auswahl sein wird. Der wahrscheinlichste Kandidat ist die zweithöchste Segmentklasse, also die Segmentklasse, die beim zweiten Durchlauf durch das in Fig. 1 beschriebene Konzept gefunden wird. Dies muß jedoch nicht immer so sein. Daher wird zunächst für die zweithöchste Segmentklasse (Segment 5 in Fig. 7), also Cluster 2 überprüft, ob diese Klasse nur ein einziges Segment oder genau zwei Segmente hat, wobei eines der beiden Segmente das erste Segment ist und das andere Segment der beiden das letzte Segment im Lied ist (Block 47).

[0094] Wird die Frage dagegen mit "Nein" beantwortet, hat die zweithöchste Segmentklasse also z. B. wenigstens drei Segmente, oder zwei Segmente, von denen eines innerhalb des Stücks und nicht am "Rand" des Stücks ist, so bleibt die zweite Segmentklasse vorerst in der Auswahl und wird fortan als "Second Cluster" bezeichnet.

[0095] Wird die Frage im Block 47 dagegen mit "Ja" beantwortet, scheidet die zweithöchste Klasse also aus (Block 48a), so wird sie durch die Segmentklasse ersetzt, die am häufigsten im gesamten Lied vorkommt (anders ausgedrückt: die am meisten Segmente beinhaltet) und nicht der höchsten Segmentklasse (Cluster 1) entspricht. Diese Segmentklasse wird fortan als "Second Cluster" bezeichnet.

[0096] "Second Cluster" muss sich, wie es nachfolgend dargelegt wird, noch mit einer dritten Segmentklasse messen (48b), welche als "Third Cluster" bezeichnet wird, um am Ende als Kandidat den Auswahlprozess zu überstehen.

[0097] Die Segmentklasse "Third Cluster" entspricht dem Cluster, welcher am häufigsten im gesamten Lied vorkommt, jedoch weder der höchsten Segmentklasse (Cluster 1) noch der Segmentklasse "Second Cluster" entspricht, sozusagen das am nächsthäufigsten (oftmals auch gleich häufig) vorkommende Cluster nach Cluster 1 und "Second Cluster".

[0098] Hinsichtlich der so genannten Bridge-Problematik wird nun für "Third Cluster" überprüft, ob es eher in die Strophe-Refrain-Auswahl gehört als "Second Cluster" oder nicht. Dies geschieht deshalb, da "Second Cluster" und "Third Cluster" oftmals gleich oft vorkommen, eins von beiden also evtl. eine Bridge oder ein anderes wiederkehrendes Zwischenteil darstellt. Um zu gewährleisten, dass die Segmentklasse von den beiden ausgewählt wird, die am ehesten der Strophe oder dem Refrain entspricht, also nicht einer Bridge oder einem anderen Zwischenstück, werden die in den Blöcken 49a, 49b, 49c dargestellten Untersuchungen durchgeführt.

[0099] Die erste Untersuchung in Block 49a lautet dahingehend, dass untersucht wird, ob jedes Segment aus ThirdCluster eine gewisse Mindestlänge hat, wobei als Schwellwert z. B. 4% der gesamten Liedlänge bevorzugt wird. Andere Werte zwischen 2% und 10% können ebenfalls zu sinnvollen Ergebnissen führen.

[0100] In einem Block 49b wird dann untersucht, ob ThirdCluster einen größeren Gesamtanteil am Lied hat als SecondCluster. Hierzu wird die Gesamtzeit aller Segmente in ThirdCluster aufaddiert und mit der entsprechend aufaddierten Gesamtzahl aller Segmente in SecondCluster verglichen, wobei dann ThirdCluster einen größeren Gesamtanteil am Lied als SecondCluster hat, wenn die aufaddierten Segmente in ThirdCluster einen größeren Wert ergeben als die aufaddierten Segmente in SecondCluster.

[0101] In dem Block 49c wird schließlich überprüft, ob der Abstand der Segmente aus ThirdCluster zu den Segmenten aus Cluster 1, also dem häufigsten Cluster konstant sind, d. h. ob eine Regelmäßigkeit in der Abfolge ersichtlich ist.

[0102] Sind alle diese drei Bedingungen mit "Ja" beantwortet, so kommt ThirdCluster in die Strophe-Refrain-Auswahl. Ist dagegen wenigstens eine dieser Bedingungen nicht erfüllt, so kommt ThirdCluster nicht in die Strophe-Refrain-Auswahl. Statt dessen kommt SecondCluster in die Strophe-Refrain-Auswahl, wie es durch einen Block 50 in Fig. 4a dargestellt ist. Damit ist die "Kandidatensuche" für die Strophe-Refrain-Auswahl beendet, und es wird der in Fig. 4b gezeigte Algorithmus gestartet, bei dem am Ende feststeht, welche Segmentklasse die Strophen umfasst, und welche Segmentklasse den Refrain umfasst.

[0103] An dieser Stelle sei darauf hingewiesen, dass die drei Bedingungen in den Blöcken 49a, 49b, 49c alternativ auch gewichtet werden könnten, so dass z. B. eine Nein-Antwort im Block 49a dann "überstimmt" wird, wenn sowohl die Abfrage im Block 49b als auch die Abfrage im Block 49c mit "Ja" beantwortet werden. Alternativ könnte auch eine Bedingung der drei Bedingungen hervorgehoben werden, so dass z. B. nur untersucht wird, ob es die Regelmäßigkeit der Abfolge zwischen der dritten Segmentklasse und der ersten Segmentklasse gibt, während die Abfragen in den Blöcken 49a und 49b nicht durchgeführt werden oder nur dann durchgeführt werden, wenn die Abfrage im Block 49c mit "Nein" beantwortet wird, jedoch z. B. ein relativ großer Gesamtanteil im Block 49b und relativ große Mindestmengen im Block 49a ermittelt werden.

[0104] Alternative Kombinationen sind ebenfalls möglich, wobei für eine Low-Level-Untersuchung auch nur die Abfrage eines der Blöcke 49a, 49b, 49c für bestimmte Implementierungen ausreichend sein wird.

[0105] Nachfolgend werden beispielhafte Implementierungen des Blocks 526 zum Durchführen einer Music-Summary dargelegt. So existieren verschiedene Möglichkeiten, was als Music Summary abgespeichert werden kann. Zwei davon werden nachfolgend beschrieben, nämlich die Möglichkeit mit dem Titel "Refrain" und die Möglichkeit mit dem Titel "Medley".

[0106] Die Refrain-Möglichkeit besteht darin, eine Version des Refrains als Summary zu wählen. Hierbei wird versucht, eine Ausführung des Refrains zu wählen, die möglichst zwischen 20 und 30 Sekunden lang ist. Ist ein Segment mit einer solchen Länge im Refrain-Cluster nicht enthalten, so wird eine Version gewählt, die eine möglichst geringe Abweichung zu einer Länge von 25 Sekunden hat. Ist der gewählte Refrain länger als 30 Sekunden, wird er bei diesem Ausführungsbeispiel über 30 Sekunden ausgeblendet und ist er kürzer als 20 Sekunden, so wird er mit dem darauf folgenden Segment auf 30 Sekunden verlängert.

[0107] Das Abspeichern eines Medleys für die zweite Möglichkeit entspricht noch eher einer tatsächlichen Zusammenfassung eines Musikstücks. Hierbei werden ein Ausschnitt der Strophe, ein Ausschnitt des Refrains und ein Ausschnitt eines dritten Segments in ihrer tatsächlichen chronologischen Reihenfolge als Medley konstruiert. Das dritte Segment wird aus einem Cluster ausgewählt, das den größten Gesamtanteil am Lied hat und nicht Strophe oder Refrain ist.

[0108] Mit folgender Priorität wird nach der geeignetsten Abfolge der Segmente gesucht:

- "drittes Segment"-Strophe-Refrain;
- Strophe-Refrain-"drittes Segment"; oder
- Strophe-"drittes Segment"-Refrain.

[0109] Die gewählten Segmente werden nicht in ihrer vollen Länge in des Medley eingebaut. Die Länge ist vorzugsweise auf feste 10 Sekunden pro Segment festgelegt, damit insgesamt wieder eine Summary von 30 Sekunden entsteht. Alternative Werte sind jedoch ebenfalls ohne weiteres realisierbar.

[0110] Vorzugsweise wird zur Rechenzeiteinsparung nach der Merkmalsextraktion im Block 502 bzw. nach dem Block 508 eine Gruppierung mehrerer Merkmalsvektoren im Block 510 durchgeführt, indem ein Mittelwert über die gruppierten Merkmalsvektoren gebildet wird. Die Gruppierung kann im nächsten-Verarbeitungsschritt, der Berechnung der Ähnlichkeitsmatrix, Rechenzeit einsparen. Zur Berechnung der Ähnlichkeitsmatrix wird zwischen allen möglichen Kombinationen von je zwei Merkmalsvektoren eine Distanz ermittelt. Daraus ergeben sich bei n Vektoren über das gesamte Stück $n \times n$ Berechnungen. Ein Gruppierungsfaktor g gibt an, wie viele aufeinander folgende Merkmalsvektoren über die Mittelwertbildung zu einem Vektor gruppiert werden. Dadurch kann die Anzahl der Berechnungen reduziert werden.

[0111] Die Gruppierung ist auch eine Art von Rauschunterdrückung, bei der kleine Änderungen in der Merkmalsausprägung von aufeinander folgenden Vektoren im Mittel kompensiert werden. Diese Eigenschaft hat eine positive Auswirkung auf das Auffinden von großen Liedstrukturen.

[0112] Das erfindungsgemäße Konzept ermöglicht es, mittels eines speziellen Music-Players durch die berechneten Segmente zu navigieren und gezielt einzelne Segmente anzuwählen, so dass ein Konsument in einem Musikladen durch beispielsweise Drücken einer bestimmten Taste oder durch Aktivieren eines bestimmten Software-Befehls ohne weiteres sofort zum Refrain eines Stücks springen kann, um festzustellen, ob der Refrain ihm gefällt, um dann, vielleicht noch eine Strophe anzuhören, damit der Konsument schließlich eine Kaufentscheidung treffen kann. Damit ist es einem Kaufinteressenten komfortabel möglich, von einem Stück genau das zu hören, was ihn besonders interessiert, während er sich z. B. das Solo oder die Bridge dann tatsächlich für den Hörgenuss zu Hause aufsparen kann.

[0113] Alternativ ist das erfindungsgemäße Konzept auch für einen Musikladen von großem Vorteil, da der Kunde gezielt und damit auch schnell Reinhören und letztendlich kaufen kann, so dass die Kunden nicht lange auf ein Reinhören warten müssen, sondern ebenfalls schnell an die Reihe kommen. Dies liegt daran, dass ein Benutzer nicht ständig hin- und herspulen muss, sondern gezielt und schnell sämtliche Informationen des Stücks erhält, die er auch haben möchte.

[0114] Ferner sei auf einen wesentlichen Vorteil des erfindungsgemäßen Konzepts hingewiesen, nämlich dass insbesondere auf Grund der Nachkorrektur der Segmentierung keine Informationen des Stücks verloren gehen. So werden zwar alle Segmente, die vorzugsweise kürzer als 6 Sekunden sind, mit dem Vorgänger- oder Nachfolger-Segment verschmolzen. Allerdings werden keine Segmente, so kurz sie auch sind, eliminiert. Dies hat den Vorteil, dass der Benutzer prinzipiell alles im Stück anhören kann, so dass ein zwar kurzes aber einem Benutzer doch sehr gut gefallendes markantes Stück, das bei einer Segmentierungs-Nachkorrektur weggefallen wäre, die tatsächlich einen Abschnitt des Stücks komplett eliminiert hätte, dennoch dem Benutzer zur Verfügung steht, damit er vielleicht gerade aufgrund des kurzen markanten Stücks eine wohl überlegte Kaufentscheidung treffen kann.

[0115] Die vorliegende Erfindung ist jedoch auch in anderen Anwendungsszenarien anwendbar, beispielsweise beim

Werbe-Monitoring, also dort, wo ein Werbekunde überprüfen möchte, ob das Audiostück, für das er Werbezeit gekauft hat, auch tatsächlich über die gesamte Länge abgespielt worden ist. Ein Audiostück kann beispielsweise Musiksegmente, Sprechersegmente, und Geräuschsegmente umfassen. Der Segmentierungsalgorithmus, also die Segmentierung und nachfolgende Einordnung in Segmentgruppen ermöglicht dann eine schnelle und wesentlich weniger aufwendige Überprüfung als ein kompletter Sample-weiser Vergleich. Die effiziente Überprüfung würde einfach in einer Segmentklassen-Statistik bestehen, also einem Vergleich, wie viel Segmentklassen gefunden worden sind, und wie viel Segmente in den einzelnen Segmentklassen sind, mit einer Vorgabe aufgrund des idealen Werbestücks. Damit ist es einem Werbekunden ohne weiteres möglich, zu erkennen, ob ein Rundfunksender oder Fernsehsender tatsächlich alle Hauptteile (Abschnitte) des Werbesignals ausgestrahlt hat oder nicht.

[0116] Die vorliegende Erfindung ist ferner dahingehend von Vorteil, dass sie zur Recherche in großen Musikdatenbanken eingesetzt werden kann, um beispielsweise nur die Refrains von vielen Musikstücken durchzuhören, um danach eine Musikprogrammauswahl durchzuführen. In diesem Fall würden nur einzelne Segmente aus der mit "Refrain" markierten Segmentklasse von vielen verschiedenen Stücken ausgewählt werden und von einem Programmanbieter bereitgestellt werden. Alternativ könnte auch ein Interesse dahingehend bestehen, von einem Interpreten alle z. B. Gitarrensolos miteinander zu vergleichen. Erfindungsgemäß können diese ebenfalls ohne weiteres bereitgestellt werden, indem immer eines oder mehrere Segmente (falls vorhanden) in der mit "Solo" bezeichneten Segmentklasse aus einer großen Anzahl von Musikstücken z. B. zusammengefügt und als ein File bereitgestellt werden.

[0117] Wieder andere Anwendungsmöglichkeiten bestehen darin, aus verschiedenen Audiostücken Strophen und Refrains zu mischen, was insbesondere für DJs von Interesse sein wird und völlig neue Möglichkeiten der kreativen Musiksynthese eröffnet, welche einfach und vor allem automatisch zielgenau durchgeführt werden können. So ist das erfindungsgemäße Konzept ohne weiteres automatisierbar, da es an keiner Stelle eine Benutzerintervention erfordert. Dies bedeutet, dass Nutzer des erfindungsgemäßen Konzepts keineswegs eine besondere Ausbildung benötigen, außer z. B. eine übliche Geschicktheit im Umgang mit normalen Software-Benutzeroberflächen.

[0118] Je nach den praktischen Gegebenheiten kann das erfindungsgemäße Konzept 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. In anderen Worten ausgedrückt stellt die Erfindung somit ein Computerprogramm mit einem Programmcode zur Durchführen des Verfahrens dar, wenn das Computerprogramm auf einem Computer abläuft.

Patentansprüche

1. Vorrichtung zum Gruppieren von zeitlichen Segmenten eines Audiostücks, das in in dem Audiostück wiederholt auftretende Hauptteile gegliedert ist, in verschiedene Segmentklassen, wobei eine Segmentklasse einem Hauptteil zugeordnet ist, mit folgenden Merkmalen:

einer Einrichtung (10) zum Bereitstellen einer Ähnlichkeitsdarstellung für die Segmente, wobei die Ähnlichkeitsdarstellung für jedes Segment eine zugeordnete Mehrzahl von Ähnlichkeitswerten aufweist, wobei die zugeordnete Mehrzahl von Ähnlichkeitswerten angibt, wie ähnlich das Segment zu jedem anderen Segment des Audiostücks ist;

einer Einrichtung (12) zum Berechnen eines Ähnlichkeits-Schwellwerts für ein spezifisches Segment ausschließlich unter Verwendung der Mehrzahl der Ähnlichkeitswerte, die dem Segment zugeordnet sind; und einer Einrichtung (14) zum Zuweisen eines Segments zu einer Segmentklasse, wenn der Ähnlichkeitswert des Segments eine vorbestimmte Bedingung bezüglich des Ähnlichkeits-Schwellwerts erfüllt.

2. Vorrichtung nach Anspruch 1, die ferner folgendes Merkmal aufweist:

eine Segmentauswahleinrichtung (16) zum Ermitteln eines Extrem-Segments, dessen zugeordnete Mehrzahl von Ähnlichkeitswerten zusammen addiert ein Extremum aufweist, wobei die Einrichtung (12) zum Berechnen ausgebildet ist, um für das Extremsegment den Ähnlichkeits-Schwellwert zu berechnen, und wobei die Einrichtung (14) zum Zuweisen ausgebildet ist, um die Segmentklasse mit einem Hinweis auf das Extremsegment zu kennzeichnen.

3. Vorrichtung nach Anspruch 1 oder 2, bei der die Einrichtung (14) zum Zuweisen ausgebildet ist, um ein Segment,

das die vorbestimmte Bedingung bezüglich des Ähnlichkeits-Schwellwerts nicht erfüllt, der Segmentklasse nicht zuzuordnen sondern für eine Zuordnung zu einer anderen Segmentklasse zu belassen, und wobei die Einrichtung (14) zum Zuweisen ausgebildet ist, um für ein zugeordnetes Segment den Ähnlichkeitswert des zugeordneten Segments bei einer Zuweisung zu einer anderen Segmentklasse nicht mehr zu berücksichtigen.

- 5
4. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (12) zum Berechnen des Ähnlichkeits-Schwellwerts in einem späteren Durchgang ausgebildet ist, um nach einer früheren Zuweisung einer Segmentklasse Ähnlichkeitswerte für früher zugewiesene Segmente in der Mehrzahl von Ähnlichkeitswerten zu ignorieren, und
- 10 bei der die Einrichtung (14) zum Zuweisen ausgebildet ist, um in einem späteren Durchgang eine Zuweisung zu einer anderen Segmentklasse als der Segmentklasse bei einem früheren Durchgang auszuführen.

5. Vorrichtung nach einem der vorhergehenden Ansprüche, die ferner folgendes Merkmal aufweist:

15 eine Segmentzuweisungskonflikteinrichtung (18), die ausgebildet ist, um in dem Fall, in dem durch die Einrichtung (14) zum Zuweisen ein Konfliktsegment zu zwei verschiedenen Segmentklassen zugeordnet werden sollte, einen ersten Ähnlichkeitswert des Konfliktsegments mit einem Segment einer ersten Segmentklasse zu ermitteln, und um einen zweiten Ähnlichkeitswert des Konfliktsegments mit einem Segment einer zweiten Segmentklasse zu ermitteln, und

20 wobei die Einrichtung (14) zum Zuweisen ausgebildet ist, um in dem Fall, in dem der zweite Ähnlichkeitswert auf eine stärkere Ähnlichkeit des Konfliktsegments mit dem Segment der zweiten Segmentklasse hinweist, das Konfliktsegment aus der ersten Segmentklasse zu entfernen und der zweiten Segmentklasse zuzuweisen.

- 25 6. Vorrichtung nach Anspruch 5, bei der die Segmentzuweisungskonflikteinrichtung (18) ausgebildet ist, um im Falle einer Entfernung des Segments aus der ersten Segmentklasse dem Segment eine Tendenz auf die erste Segmentklasse zuzuweisen, oder um im Falle einer nicht erfolgten Entfernung des Segments dem Segment eine Tendenz auf die zweite Segmentklasse zuzuweisen.

- 30 7. Vorrichtung nach einem der vorhergehenden Ansprüche, die ferner folgendes Merkmal aufweist:

eine Segmentierungskorrektureinrichtung (20), die ausgebildet ist, um eine Segmentierung des Audiostücks zu korrigieren, wobei die Segmentierungskorrektureinrichtung (20) ausgebildet ist, um Segmente abhängig von Segmentklasseninformationen für die Segmente mit einem vorangehenden Segment oder einem nachfolgenden Segment zu verschmelzen.

- 35 8. Vorrichtung nach Anspruch 7, bei der die Segmentierungskorrektureinrichtung (20) ausgebildet ist, um für ein Segment, das kürzer als eine vorbestimmte Minimallänge ist, festzustellen, ob eine Tendenz des Segments mit einer Segmentklasse übereinstimmt, der ein unmittelbar zeitlich vorausgehendes Segment angehört, und um in diesem Fall das Segment mit dem zeitlich unmittelbar vorausgehenden Segment zu verschmelzen, oder die ausgebildet ist, um für ein Segment, das kürzer als eine vorbestimmte Minimallänge ist, festzustellen, ob eine Tendenz des Segments auf eine Segmentklasse hinweist, der ein zeitlich unmittelbar nachfolgendes Segment angehört, und um in diesem Fall das Segment mit dem zeitlich unmittelbar nachfolgenden Segment zu verschmelzen.
- 40

- 45 9. Vorrichtung nach einem der vorhergehenden Ansprüche, die eine Segmentierungskorrektureinrichtung (20) aufweist, die ausgebildet ist, um zeitlich aufeinander folgende Segmente, die der gleichen Segmentklasse angehören, zu verschmelzen.

- 50 10. Vorrichtung nach einem der Ansprüche 7 bis 9, bei der die Segmentierungskorrektureinrichtung (20) ausgebildet ist, um zum Korrigieren der Segmente lediglich Segmente auszuwählen, die eine zeitliche Segmentlänge haben, die kürzer als eine vorbestimmte Minimallänge ist.

- 55 11. Vorrichtung nach Anspruch 10, bei der die Segmentierungskorrektureinrichtung (20) ausgebildet ist, um ein ausgewähltes Segment aus einer zweiten Segmentklasse, dessen zeitliches Vorläufersegment, und dessen zeitliches Nachfolgersegment einer ersten Segmentklasse angehören, mit dem Vorgänger-Segment und dem Nachfolger-Segment zu verschmelzen.

12. Vorrichtung nach Anspruch 10 oder 11, bei der die Segmentierungskorrektureinrichtung (20) ausgebildet ist, um ein Segment, das in einer Segmentklasse ist, die nur ein einziges Segment umfasst, mit dem vorausgehenden

Segment oder dem nachfolgenden Segment zu verschmelzen.

13. Vorrichtung nach Anspruch 10, 11 oder 12, bei der die Segmentierungskorrekturereinrichtung (20) ausgebildet ist, um mehrere ausgewählte Segmente, die in derselben Segmentklasse sind, mit jeweils einem zeitlich vorausgehenden Segment oder jeweils einem zeitlich nachfolgenden Segment zu verschmelzen, wenn sämtliche ausgewählten Segmente der Segmentklasse Vorläufer-Segmente aus ein und derselben Segmentklasse oder Nachfolger-Segmente aus ein und derselben Segmentklasse umfassen.
14. Vorrichtung nach einem der Ansprüche 7 bis 13, bei der die Segmentierungskorrekturereinrichtung ausgebildet ist, um für ein Segment, das eine kleinere zeitliche Länge als eine vorbestimmte Minimallänge aufweist, einen ersten Neuheitswert an einem Anfang des Segments zu ermitteln, und einen zweiten Neuheitswert an einem Ende des Segments zu ermitteln, und um das Segment mit einem zeitlich folgenden Segment zu verschmelzen, wenn der erste Neuheitswert größer als der zweite Neuheitswert ist, oder um das Segment mit einem zeitlich vorhergehenden Segment zu verschmelzen, wenn der erste Neuheitswert kleiner als der zweite Neuheitswert ist.
15. Vorrichtung nach einem der Ansprüche 7 bis 14, bei der die Segmentierungskorrekturereinrichtung (20) ausgebildet ist, um abhängig von verschiedenen vorbestimmten Segmentlängen verschiedene Korrekturmaßnahmen durchzuführen.
16. Vorrichtung nach einem der vorhergehenden Ansprüche, die ferner eine Segmentklassen-Bezeichnungseinrichtung aufweist, die ausgebildet ist, um abhängig von einer zeitlichen Position von Segmenten in verschiedenen Segmentklassen eine Bezeichnung von Segmentklassen zu unterschiedlichen Hauptteilen durchzuführen.
17. Vorrichtung nach Anspruch 16, bei der die Segmentklassen-Bezeichnungseinrichtung (22) ausgebildet ist, um vor einer Segmentklassen-Bezeichnung in einen Hauptteil "Strophe" und in einen Hauptteil "Refrain" zwei Segmentklassen-Kandidaten zur Berücksichtigung der Segmente in den Segmentklassen auszuwählen.
18. Vorrichtung nach Anspruch 16 oder 17, bei der die Segmentklassen-Bezeichnungseinrichtung (22) ausgebildet ist, um eine Kandidaten-Segmentklasse als Refrain-Klasse zu bezeichnen, wenn die Kandidaten-Segmentklasse das Segment umfasst, das in dem Audiostück zeitlich nach allen anderen Segmenten der anderen Kandidaten-Segmentklasse vorkommt.
19. Vorrichtung nach einem der Anspruch 16 bis 18, bei der Segmentklassen-Bezeichnungseinrichtung (22) ausgebildet ist, um eine Kandidaten-Segmentklasse als Strophe-Klasse zu bezeichnen, wenn die Kandidaten-Segmentklasse nicht das Segment umfasst, das in dem Audiostück zeitlich nach allen anderen Segmenten der anderen Kandidaten-Segmentklasse vorkommt.
20. Verfahren zum Gruppieren von zeitlichen Segmenten eines Audiostücks, das in in dem Audiostück wiederholt auftretende Hauptteile gegliedert ist, in verschiedene Segmentklassen, wobei eine Segmentklasse einem Hauptteil zugeordnet ist, mit folgenden Schritten:

Bereitstellen (10) einer Ähnlichkeitsdarstellung für die Segmente, wobei die Ähnlichkeitsdarstellung für jedes Segment eine zugeordnete Mehrzahl von Ähnlichkeitswerten aufweist, wobei die zugeordnete Mehrzahl von Ähnlichkeitswerten angibt, wie ähnlich das Segment zu jedem anderen Segment des Audiostücks ist;
Berechnen (12) eines Ähnlichkeits-Schwellwerts für ein spezifisches Segment ausschließlich unter Verwendung der Mehrzahl der Ähnlichkeitswerte, die dem Segment zugeordnet sind; und
Zuweisen (14) eines Segments zu einer Segmentklasse, wenn der Ähnlichkeitswert des Segments eine vorbestimmte Bedingung bezüglich des Ähnlichkeits-Schwellwerts erfüllt.
21. Computer-Programm mit einem Programmcode zum Ausführen des Verfahrens nach Patentanspruch 20, wenn das Computer-Programm auf einem Rechner läuft.

Claims

1. Apparatus for grouping temporal segments of an audio track, which is subdivided into main parts occurring repeatedly in the audio track, into different segment classes, a segment class being associated with a main part, having the following features:

a device (10) for providing a similarity representation for the segments, the similarity representation for each segment having an associated plurality of similarity values, the associated plurality of similarity values indicating how similar the segment is to each other segment of the audio track;

a device (12) for calculating a similarity threshold value for a specific segment solely using the plurality of similarity values associated with the segment; and

a device (14) for allocating a segment to a segment class if the similarity value of the segment meets a predetermined condition with respect to the similarity threshold value.

2. Apparatus according to Claim 1, which also has the following feature:

a segment selection device (16) for determining an extreme segment whose associated plurality of similarity values, added together, has an extremum,
the calculating device (12) being designed to calculate the similarity threshold value for the extreme segment, and
the allocating device (14) being designed to label the segment class with an indication of the extreme segment.

3. Apparatus according to Claim 1 or 2, in which the allocating device (14) is designed not to associate a segment, which does not meet the predetermined condition with respect to the similarity threshold value, with the segment class, but rather to leave it for association with a different segment class, and
the allocating device (14) being designed to no longer take into account, for an associated segment, the similarity value of the associated segment during allocation to a different segment class.

4. Apparatus according to one of the preceding claims, in which the device (12) for calculating the similarity threshold value is designed, in a subsequent pass, following previous allocation to a segment class, to ignore similarity values for previously allocated segments in the plurality of similarity values, and
in which the allocating device (14) is designed, in a subsequent pass, to carry out allocation to a segment class other than the segment class in a previous pass.

5. Apparatus according to one of the preceding claims, which also has the following feature:

a segment allocation conflict device (18) which is designed, in the case in which a conflict segment should be associated with two different segment classes by the allocating device (14), to determine a first similarity value of the conflict segment to a segment in a first segment class and to determine a second similarity value of the conflict segment to a segment in a second segment class, and
the allocating device (14) being designed, in the case in which the second similarity value indicates stronger similarity of the conflict segment to the segment in the second segment class, to remove the conflict segment from the first segment class and to allocate it to the second segment class.

6. Apparatus according to Claim 5, in which the segment allocation conflict device (18) is designed, if the segment is removed from the first segment class, to allocate a tendency towards the first segment class to the segment or, if the segment has not been removed, to allocate a tendency towards the second segment class to the segment.

7. Apparatus according to one of the preceding claims, which also has the following feature:

a segmentation correction device (20) which is designed to correct segmentation of the audio track, the segmentation correction device (20) being designed to merge segments with a preceding segment or a subsequent segment depending on segment class information for the segments.

8. Apparatus according to Claim 7, in which the segmentation correction device (20) is designed to determine, for a segment which is shorter than a predetermined minimum length, whether a tendency of the segment corresponds to a segment class to which an immediately temporally preceding segment belongs, and, in this case, to merge the segment with the temporally immediately preceding segment, or is designed to determine, for a segment which is shorter than a predetermined minimum length, whether a tendency of the segment indicates a segment class to which a temporally immediately subsequent segment belongs, and, in this case, to merge the segment with the temporally immediately subsequent segment.

9. Apparatus according to one of the preceding claims, which has a segmentation correction device (20) designed to merge temporally successive segments belonging to the same segment class.

10. Apparatus according to one of Claims 7 to 9, in which the segmentation correction device (20), in order to correct the segments, is designed to select only segments having a temporal segment length which is shorter than a predetermined minimum length.
- 5 11. Apparatus according to Claim 10, in which the segmentation correction device (20) is designed to merge a selected segment from a second segment class, the temporal precursor segment of which and the temporal successor segment of which belong to a first segment class, with the precursor segment and the successor segment.
- 10 12. Apparatus according to Claim 10 or 11, in which the segmentation correction device (20) is designed to merge a segment, which is in a segment class comprising only a single segment, with the preceding segment or the subsequent segment.
- 15 13. Apparatus according to Claim 10, 11 or 12, in which the segmentation correction device (20) is designed to merge a plurality of selected segments, which are in the same segment class, with a temporally preceding segment or a temporally subsequent segment in each case if all selected segments in the segment class comprise precursor segments from the same segment class or successor segments from the same segment class.
- 20 14. Apparatus according to one of Claims 7 to 13, in which the segmentation correction device is designed to determine, for a segment having a shorter temporal length than a predetermined minimum length, a first novelty value at a start of the segment and to determine a second novelty value at an end of the segment and to merge the segment with a temporally subsequent segment if the first novelty value is greater than the second novelty value or to merge the segment with a temporally preceding segment if the first novelty value is less than the second novelty value.
- 25 15. Apparatus according to one of Claims 7 to 14, in which the segmentation correction device (20) is designed to carry out different corrective measures on the basis of different predetermined segment lengths.
- 30 16. Apparatus according to one of the preceding claims, which also has a segment class naming device which is designed to name segment classes for different main parts depending on a temporal position of segments in different segment classes.
- 35 17. Apparatus according to Claim 16, in which the segment class naming device (22) is designed, before naming segment classes as a "verse" main part and a "refrain" main part, to select two segment class candidates for taking into account the segments in the segment classes.
- 40 18. Apparatus according to Claim 16 or 17, in which the segment class naming device (22) is designed to name a candidate segment class as a refrain class if the candidate segment class comprises the segment which temporally occurs after all other segments in the other candidate segment class in the audio track.
- 45 19. Apparatus according to one of Claims 16 to 18, in which the segment class naming device (22) is designed to name a candidate segment class as a verse class if the candidate segment class does not comprise the segment which temporally occurs after all other segments in the other candidate segment class in the audio track.
20. Method for grouping temporal segments of an audio track, which is subdivided into main parts occurring repeatedly in the audio track, into different segment classes, a segment class being associated with a main part, having the following steps of:
 - providing (10) a similarity representation for the segments, the similarity representation for each segment having an associated plurality of similarity values, the associated plurality of similarity values indicating how similar the segment is to each other segment of the audio track;
 - 50 calculating (12) a similarity threshold value for a specific segment solely using the plurality of similarity values associated with the segment; and
 - allocating (14) a segment to a segment class if the similarity value of the segment meets a predetermined condition with respect to the similarity threshold value.
- 55 21. Computer program having a program code for carrying out the method according to Patent Claim 20 when the computer program runs on a computer.

Revendications

1. Ensemble de regroupement en différentes classes de segments de segments temporels d'un morceau audio divisé en parties principales qui reviennent de manière répétée dans le morceau audio, une classe de segments étant associée à une partie principale, l'ensemble présentant les caractéristiques suivantes :

un dispositif (10) qui prépare une représentation de similitude pour les segments, la représentation de similitude comprenant pour chaque segment une pluralité associée de valeurs de similitude, la pluralité associée de valeurs de similitude indiquant dans quelle mesure le segment est similaire à chaque autre segment du morceau audio,
 un dispositif (12) qui calcule une valeur de seuil de similitude pour un segment spécifique en recourant exclusivement à la pluralité de valeurs de similitude associées aux segments et
 un dispositif (14) qui attribue un segment à une classe de segments si la valeur de similitude du segment remplit une condition prédéterminée par rapport à la valeur de seuil de similitude.

2. Ensemble selon la revendication 1, qui présente en outre les caractéristiques suivantes :

un dispositif (16) de sélection de segment qui détermine un segment extrême dont la pluralité associée de valeurs de similitude s'additionnent pour former un extrémum,
 le dispositif de calcul (12) étant configuré pour calculer la valeur de seuil de similitude pour le segment extrême et le dispositif d'attribution (14) étant configuré pour caractériser la classe de segments par une indication du segment extrême.

3. Ensemble selon la revendication 1 ou 2, dans lequel le dispositif d'attribution (14) est configuré pour ne pas attribuer à la classe des segments un segment qui ne remplit pas la condition prédéterminée par rapport à la valeur de seuil de similitude mais pour l'attribuer à une autre classe de segments,
 le dispositif d'attribution (14) étant configuré pour ne plus tenir compte, pour un segment associé, de la valeur de similitude du segment associé lorsqu'il a été attribué à une autre classe de segments.

4. Ensemble selon l'une des revendications précédentes, dans lequel le dispositif (12) qui calcule la valeur de seuil de similitude est configuré pour, après une attribution antérieure à une classe de segments, ignorer dans un passage ultérieur parmi la pluralité de valeurs de similitude les valeurs de similitude de segments attribués antérieurement, le dispositif d'attribution (14) étant configuré pour, lors d'un passage ultérieur, exécuter l'attribution à une autre classe de segments que la classe de segments d'un passage antérieur.

5. Ensemble selon l'une des revendications précédentes, qui présente en outre les caractéristiques suivantes :

un dispositif (18) de gestion des conflits d'attribution de segments configuré pour, au cas où le dispositif d'attribution (14) devrait associer un segment conflictuel à deux classes de segments différentes, déterminer une première valeur de similitude du segment conflictuel par rapport à un segment d'une première classe de segments et déterminer une deuxième valeur de similitude du segment conflictuel par rapport à un segment d'une deuxième classe de segments, et
 le dispositif d'attribution (14) étant configuré pour, au cas où la deuxième valeur de similitude indique une similitude plus forte du segment conflictuel par rapport au segment de la deuxième classe de segments, enlever le segment conflictuel de la première classe de segments et l'attribuer à la deuxième classe de segments.

6. Ensemble selon la revendication 5, dans lequel le dispositif (18) de gestion des conflits d'attribution de segments est configuré pour attribuer au segment une tendance vers la première classe de segments au cas où le segment est enlevé de la première classe de segments, ou pour attribuer au segment une tendance vers la deuxième classe de segments au cas où le segment n'a pas été enlevé.

7. Ensemble selon l'une des revendications précédentes, qui présente en outre la caractéristique suivante :

un dispositif (20) de correction de la segmentation configuré pour corriger la segmentation du morceau audio, le dispositif (20) de correction de la segmentation étant configuré pour fondre des segments avec un segment précédent ou un segment suivant en fonction d'informations de classification des segments.

8. Ensemble selon la revendication 7, dans lequel le dispositif (20) de correction de la segmentation est configuré pour

vérifier sur un segment plus court qu'une longueur minimale prédéterminée s'il présente une tendance vers une classe de segments à laquelle appartient un segment immédiatement précédent et pour, dans ce cas, fondre le segment avec le segment qui précède immédiatement, ou qui est configuré pour vérifier sur un segment plus court qu'une longueur minimale prédéterminée si le segment présente une tendance vers une classe de segments à laquelle appartient un segment immédiatement suivant et pour, dans ce cas, fondre le segment avec le segment immédiatement suivant.

9. Ensemble selon l'une des revendications précédentes, qui présente un dispositif (20) de correction de la segmentation configuré pour fondre des segments successifs qui appartiennent à la même classe de segments.

10. Ensemble selon l'une des revendications 7 à 9, dans lequel le dispositif (20) de correction de la segmentation est configuré pour sélectionner pour la correction de segments uniquement les segments qui ont une longueur temporelle plus courte qu'une longueur minimale prédéterminée.

11. Ensemble selon la revendication 10, dans lequel le dispositif (20) de correction de la segmentation est configuré pour fondre un segment sélectionné provenant d'une deuxième classe de segments et dont le segment précédent et le segment suivant appartiennent à une première classe de segments avec le segment précédent et le segment suivant.

12. Ensemble selon la revendication 10 ou 11, dans lequel le dispositif (20) de correction de la segmentation est configuré pour fondre avec le segment précédent ou le segment suivant un segment qui est dans une classe de segments qui ne comporte qu'un seul segment.

13. Ensemble selon la revendication 10, 11 ou 12, dans lequel le dispositif (20) de correction de la segmentation est configuré pour fondre avec un segment précédent respectif ou un segment suivant respectif plusieurs segments sélectionnés situés dans la même classe de segments lorsque tous les segments sélectionnés de la classe de segments comprennent des segments précédents d'une seule et même classe de segments ou des segments suivants d'une seule et même classe de segments.

14. Ensemble selon l'une des revendications 7 à 13, dans lequel le dispositif de correction de la segmentation est configuré pour déterminer une première valeur de nouveauté au début d'un segment dont la longueur temporelle est inférieure à une longueur minimale prédéterminée, pour déterminer une deuxième valeur de nouveauté à une extrémité du segment et pour fondre le segment avec un segment suivant lorsque la première valeur de nouveauté est supérieure à la deuxième valeur de nouveauté ou pour fondre le segment avec un segment précédent lorsque la première valeur de nouveauté est inférieure à la deuxième valeur de nouveauté.

15. Ensemble selon l'une des revendications 7 à 14, dans lequel le dispositif (20) de correction de la segmentation est configuré pour exécuter différentes mesures de correction en fonction de différentes longueurs prédéterminées de segments.

16. Ensemble selon l'une des revendications précédentes, qui présente en outre un dispositif de désignation des classes de segments qui est configuré pour exécuter une désignation de classes de segments en différentes parties principales en fonction d'une position temporelle de segments de différentes classes de segments.

17. Ensemble selon la revendication 16, dans lequel le dispositif (22) de désignation de classe de segments est configuré pour sélectionner deux candidats de classes de segments pour tenir compte des segments dans les classes de segments avant une désignation de la classe de segments dans une partie principale "strophe" et dans une partie principale "refrain".

18. Ensemble selon la revendication 16 ou 17, dans lequel le dispositif (22) de désignation des classes de segments est configuré pour désigner par classe de refrain un candidat de classe de segments lorsque le candidat de classe de segments comprend le segment qui survient dans le morceau audio après tous les autres segments de l'autre candidat de classe de segments.

19. Ensemble selon l'une des revendications 16 à 18, dans lequel le dispositif (22) de désignation de classe de segments est configuré pour désigner par classe de strophe un candidat de classe de segments lorsque le candidat de classe de segments ne comprend pas le segment qui survient dans le morceau audio après tous les autres segments de l'autre candidat de classe de segments.

- 20.** Procédé de regroupement en différentes classes de segments de segments temporels d'un morceau audio divisé en parties principales qui reviennent de manière répétée dans le morceau audio, une classe de segments étant associée à une partie principale, le procédé comportant les étapes suivantes :

5 préparer (10) une représentation de similitude pour les segments, la représentation de similitude comprenant pour chaque segment une pluralité associée de valeurs de similitude, la pluralité associée de valeurs de similitude indiquant dans quelle mesure le segment est similaire à chaque autre segment du morceau audio,
calculer (12) une valeur de seuil de similitude pour un segment spécifique en recourant exclusivement à la pluralité de valeurs de similitude associées aux segments et
10 attribuer (14) un segment à une classe de segments si la valeur de similitude du segment remplit une condition prédéterminée par rapport à la valeur de seuil de similitude.

- 21.** Programme informatique doté d'un code de programme permettant d'exécuter le procédé selon la revendication 20 lorsque le programme informatique est exécuté sur un ordinateur.

15

20

25

30

35

40

45

50

55

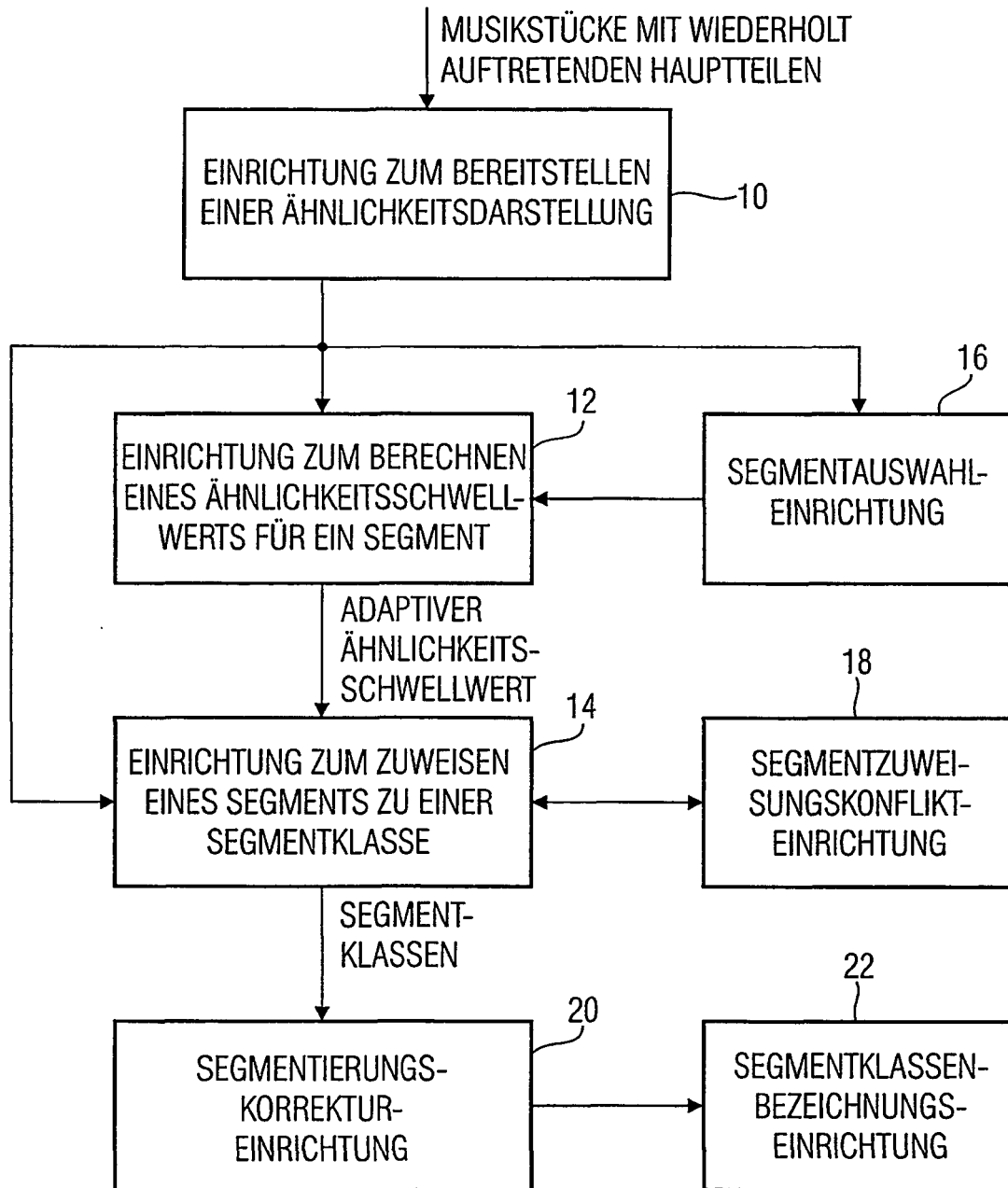


FIG. 1

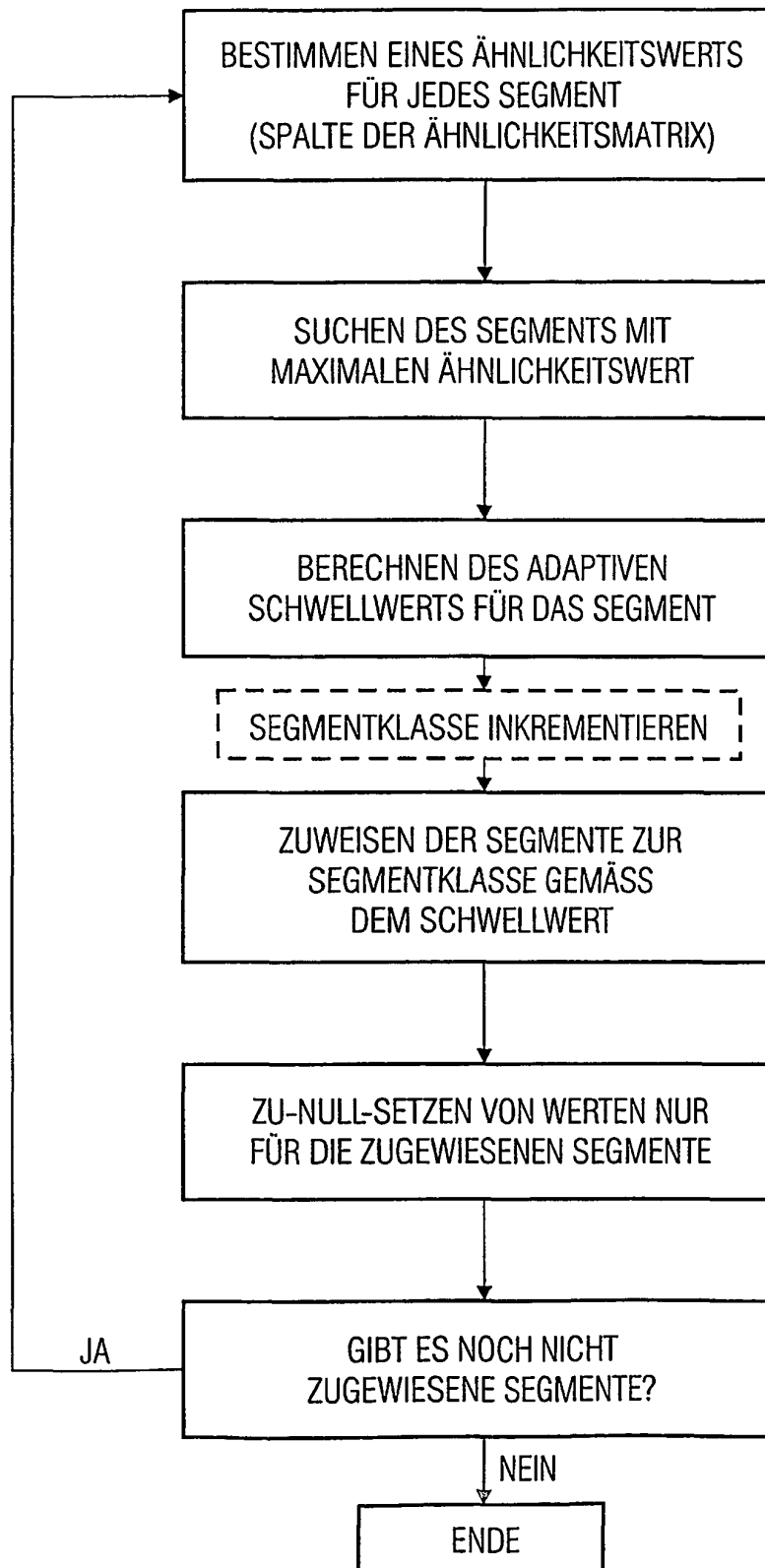


FIG. 2

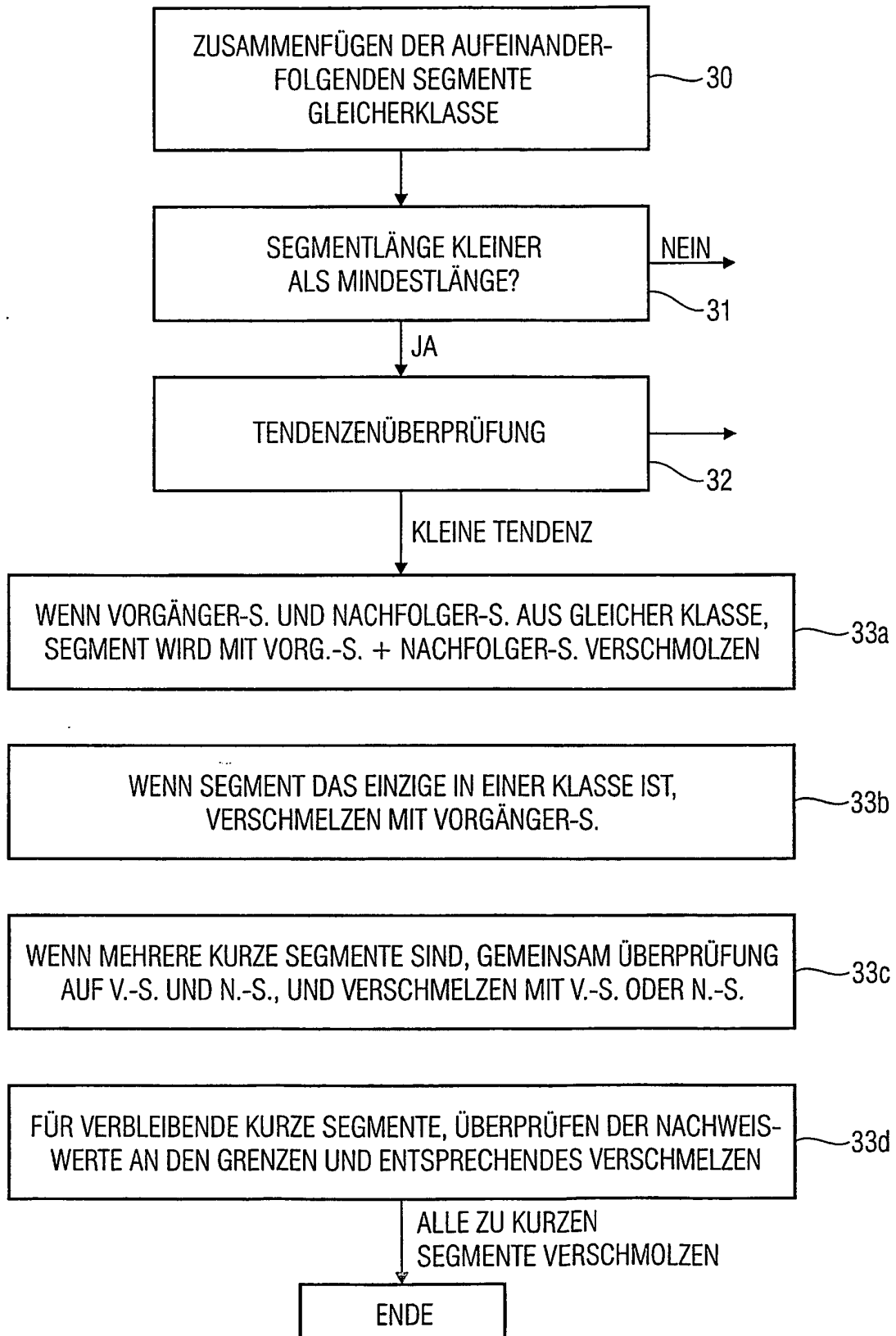


FIG. 3

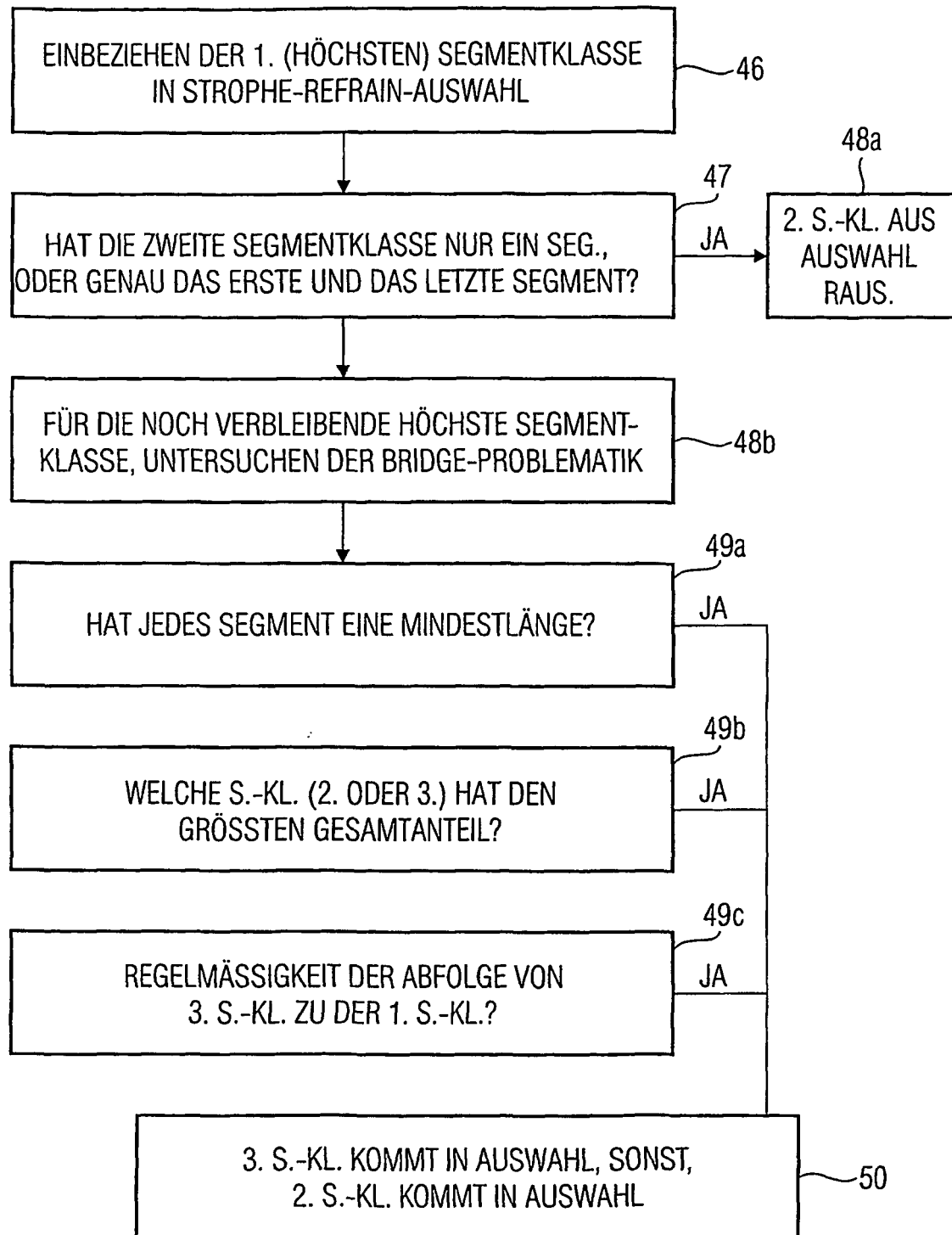


FIG. 4a

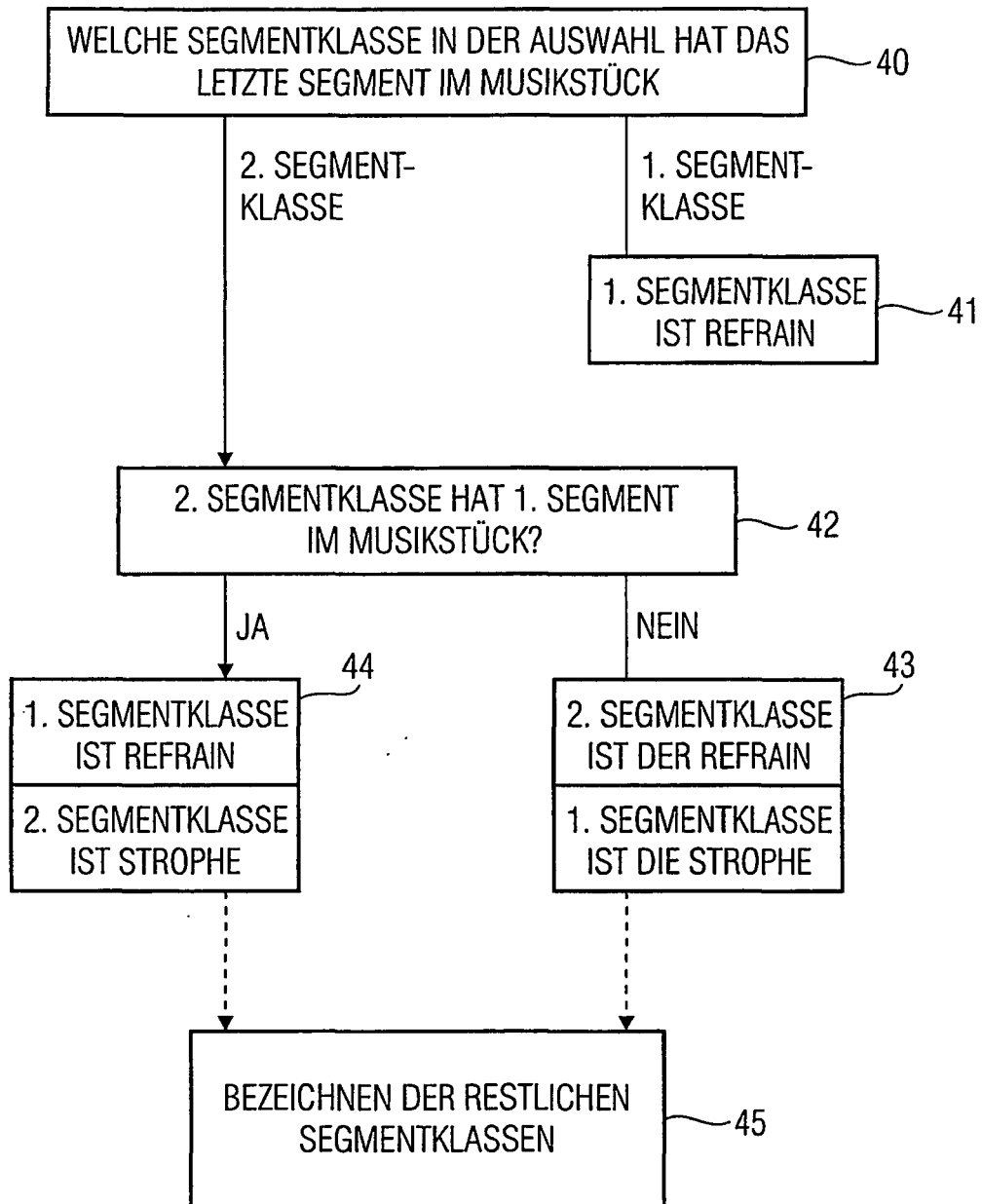
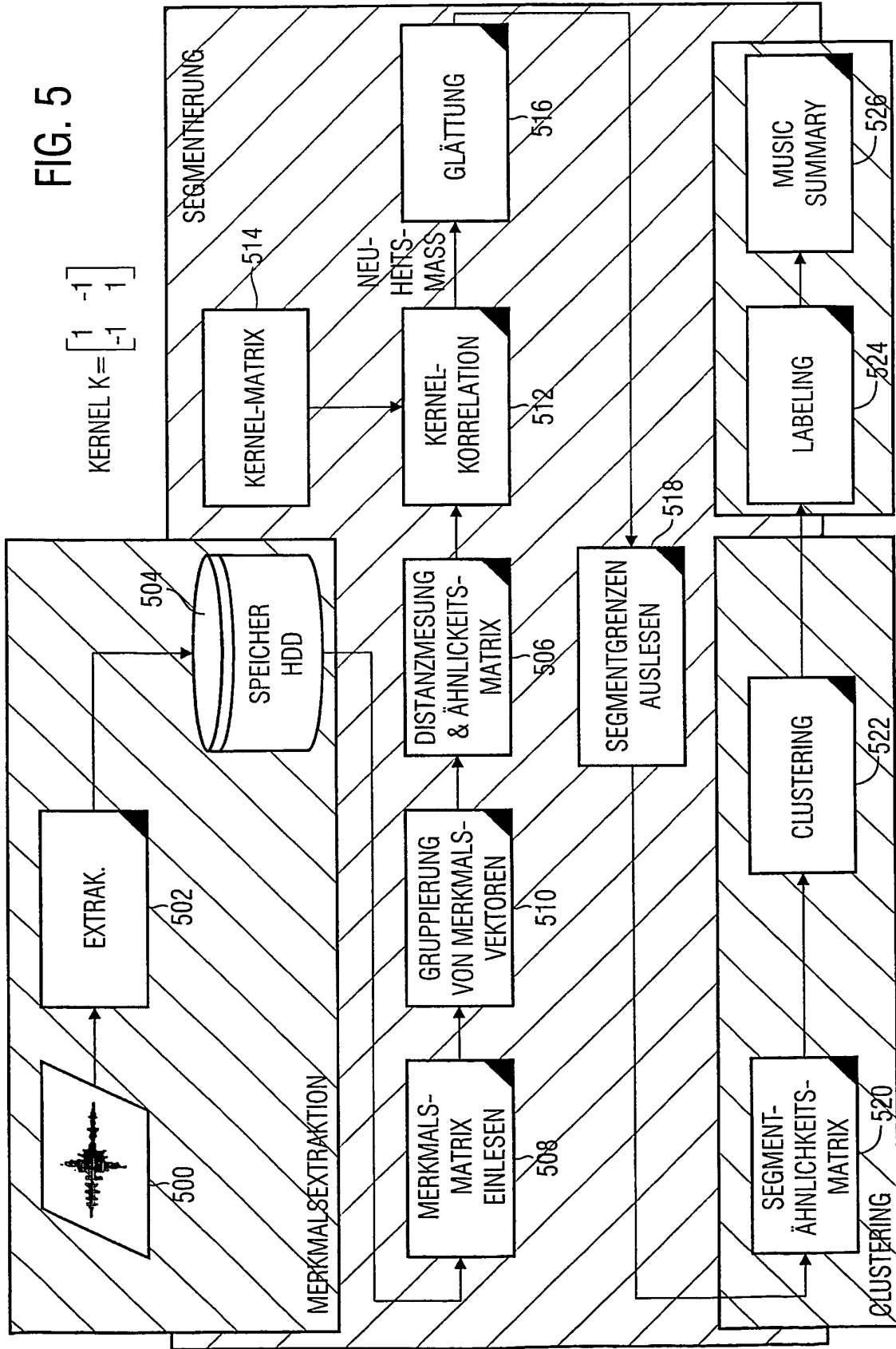


FIG. 4b

FIG. 5



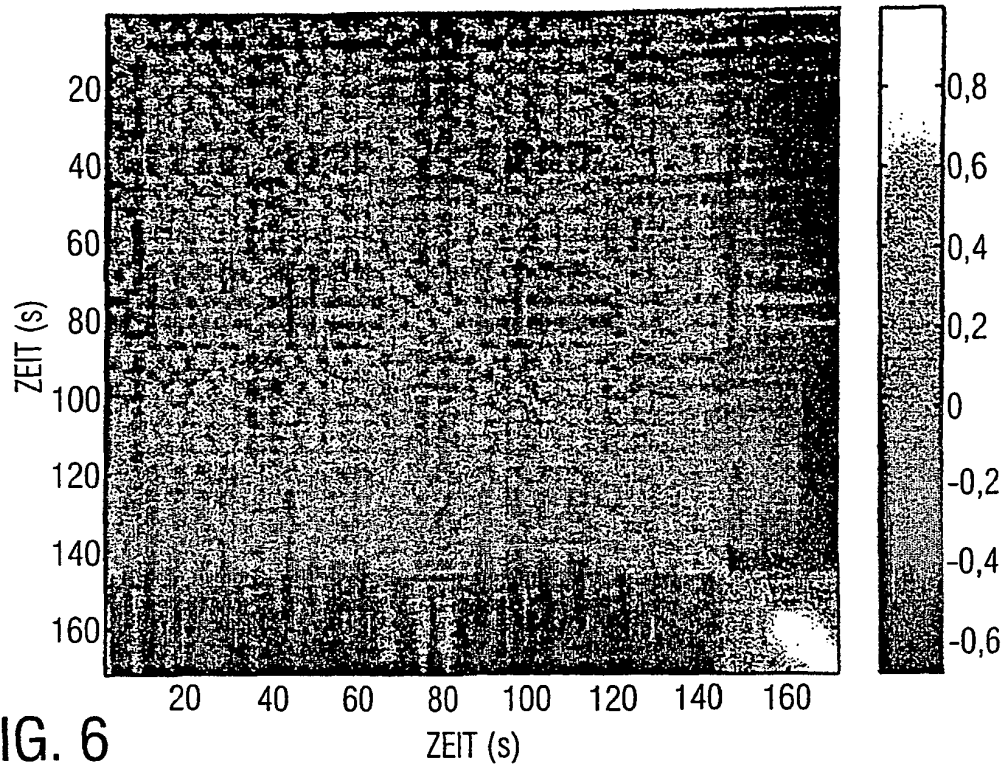


FIG. 6

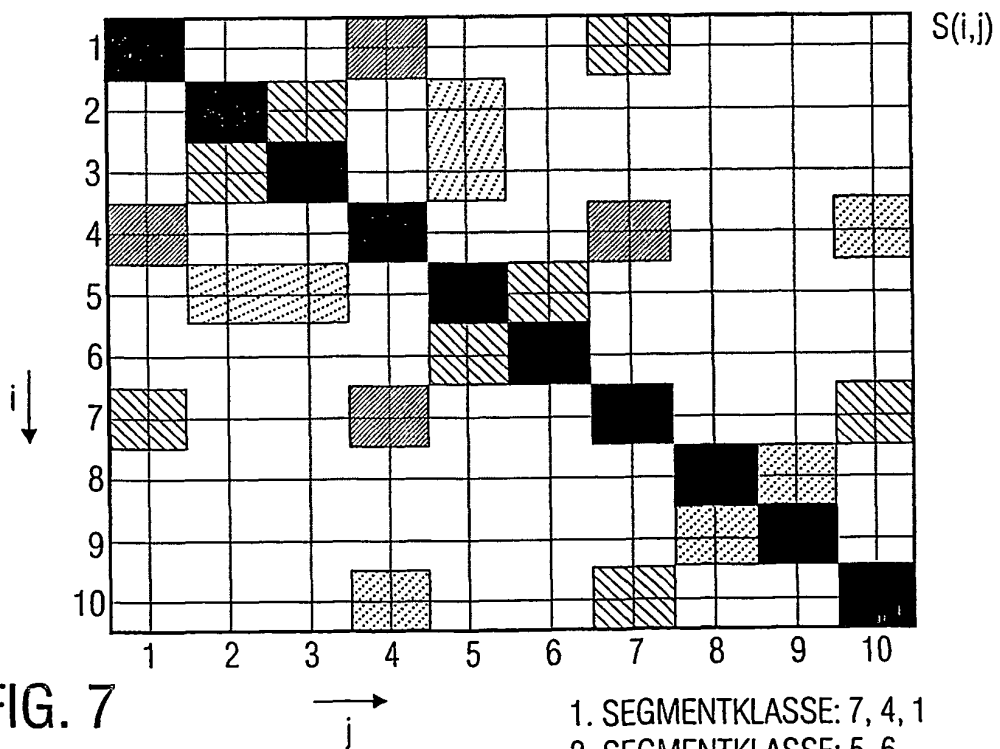


FIG. 7

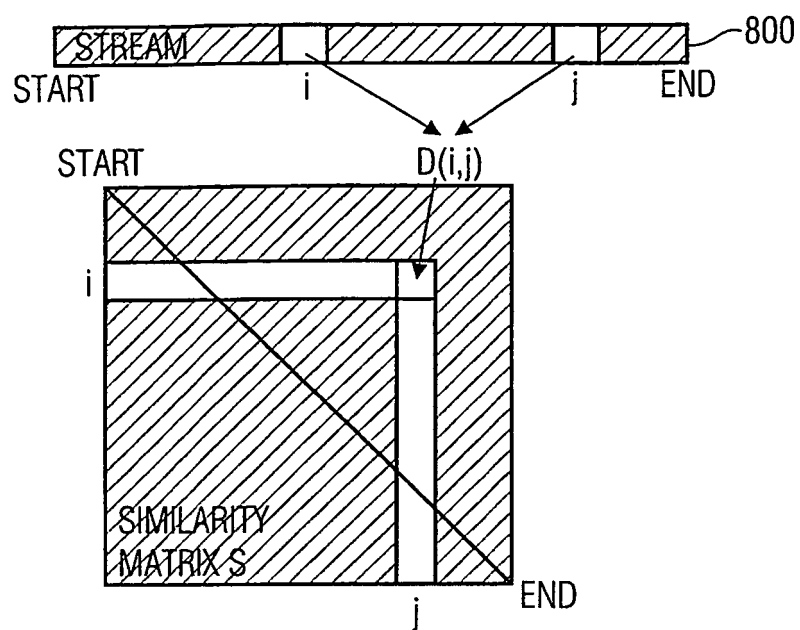


FIG. 8

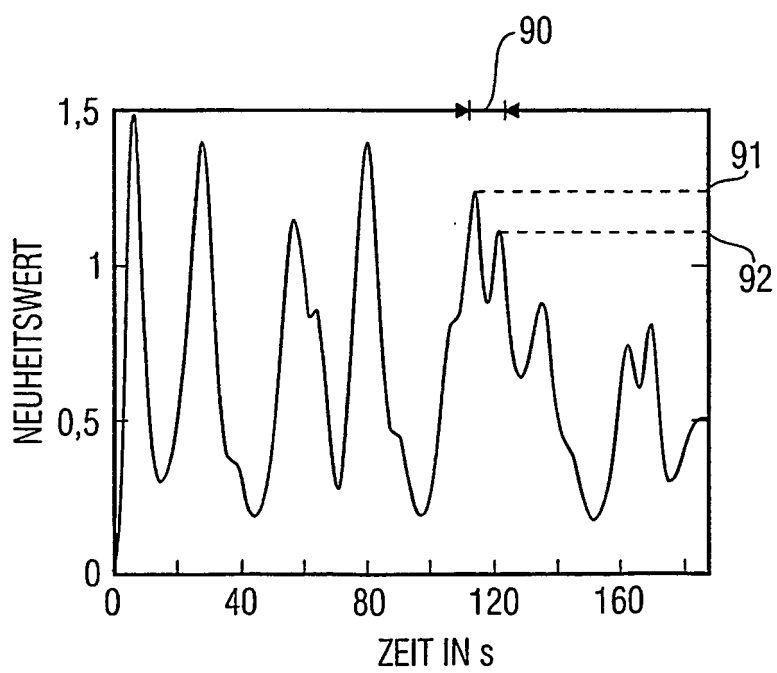


FIG. 9

IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

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

In der Beschreibung aufgeführte Patentedokumente

- EP 1577877 A1 [0033]

In der Beschreibung aufgeführte Nicht-Patentliteratur

- **FOOTE, J.T. ; COOPER, M.L.** Summarizing Popular Music via Structural Similarity Analysis. *Proceedings of the IEEE Workshop of Signal Processing to Audio and Acoustics*, 2003 [0005]
- **FOOTE, J.T. ; COOPER, M.L.** Media Segmentation using Self-Similar Decomposition. *Proceedings of SPIE Storage and Retrieval for Multimedia Databases*, Januar 2003, vol. 5021, 167-75 [0005]
- Music Summary using Key Phrases. **CHU, S. ; LOGAN B.** Technical Report. Cambridge Research Laboratory, 2000 [0028]
- **BARTSCH, M.A. ; WAKEFIELD, G. H.** To Catch a Chorus: Using Chroma-Based Representation for Audio Thumbnailing. *Proceedings of the IEEE Workshop of Signal Processing to Audio and Acoustics*, 2001, http://musen.engin.umich.edu/papers/bartsch_wakefield_waspaa01_final.pdf [0028]