



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 10 2004 023 436 B4 2006.06.14**

(12)

Patentschrift

(21) Aktenzeichen: **10 2004 023 436.1**
 (22) Anmeldetag: **10.05.2004**
 (43) Offenlegungstag: **08.12.2005**
 (45) Veröffentlichungstag
 der Patenterteilung: **14.06.2006**

(51) Int Cl.⁸: **H04N 17/00 (2006.01)**
H04L 12/26 (2006.01)
H04L 12/22 (2006.01)
H04L 9/32 (2006.01)

Innerhalb von drei Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten (§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 2 Patentkostengesetz).

(73) Patentinhaber:
m2any GmbH, 85748 Garching, DE

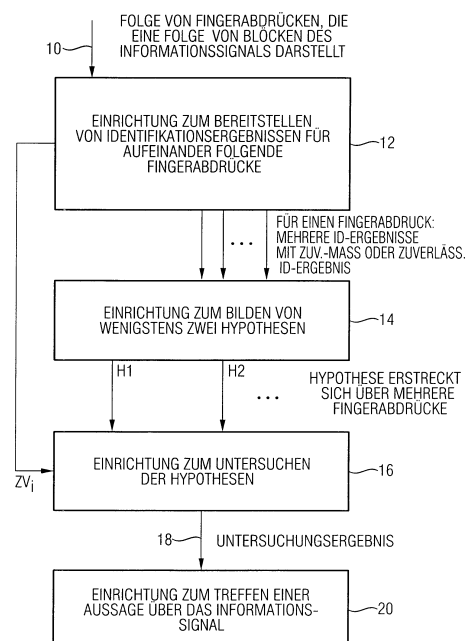
(74) Vertreter:
Schoppe, Zimmermann, Stöckeler & Zinkler, 82049 Pullach

(72) Erfinder:
Herre, Jürgen, 91054 Buckenhof, DE; Allamanche, Eric, 90425 Nürnberg, DE; Hellmuth, Oliver, 91058 Erlangen, DE; Kastner, Thorsten, 96342 Stockheim, DE

(56) Für die Beurteilung der Patentfähigkeit in Betracht gezogene Druckschriften:
DE 101 29 635 A1
WO 02/11 123 A2
WO 01/04 870 A1
A.Wang, An Industrial Strength Audio Search Algorithm, Presentation of International Symposium on Music Information Retrieval, 26-30 October 2003, Baltimore, Maryland (USA) October 2003. <http://ismir2003.ismir.net/presentations.html>;
J.Haitsma, u.a.: Robust Audio Hashing for Content Identification, In: Proceedings of the International Workshop on Content-Based Multimedia Indexing, Brescia, Italy, Sept.2001. <http://citeseer.ist.psu.edu/haitsma01robust.html>;

(54) Bezeichnung: **Vorrichtung und Verfahren zum Analysieren eines Informationssignals**

(57) Hauptanspruch: Vorrichtung zum Analysieren eines Informationssignals, das eine Folge (802) von Blöcken von Informationseinheiten (804) aufweist, wobei eine Mehrzahl von aufeinanderfolgenden Blöcken der Folge von Blöcken eine Informationsentität (806) darstellt, unter Verwendung einer Folge von Fingerabdrücken (FAi) für die Folge von Blöcken, so daß die Folge von Blöcken durch die Folge von Fingerabdrücken dargestellt ist, mit folgenden Merkmalen: einer Einrichtung (12) zum Bereitstellen von Identifikationsergebnissen (IDi) für aufeinanderfolgende Fingerabdrücke, wobei ein Identifikationsergebnis eine Zugehörigkeit eines Blocks von Informationseinheiten zu einer vorbestimmten Informationsentität darstellt, und wobei es für jedes Identifikationsergebnis ein Zuverlässigkeitsmaß gibt, wobei die Einrichtung (12) zum Bereitstellen ausgebildet ist, um für einen ersten Fingerabdruck ein erstes Identifikationsergebnis zu erzeugen, und um für einen folgenden Block ein zweites Identifikationsergebnis zu erzeugen, das sich von dem ersten Identifikationsergebnis unterscheidet; einer Einrichtung (14) zum Bilden von wenigstens zwei Hypothesen aus den Identifikationsergebnissen für die aufeinanderfolgenden Fingerabdrücke, wobei eine erste Hypothese eine...



Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf die Signalanalyse und insbesondere auf die Signalanalyse zu Zwecken der Identifikation eines Signalinhalts.

Stand der Technik

[0002] Um die immer größer werdenden Bestände an Audio- und Video-Material zu archivieren, leicht durchsuchbare Datenbanken zu erstellen oder über verschiedene Vertriebswege zu verteilen, werden automatische Informationserkennungssysteme benötigt, die helfen, Audio- und Videomaterial oder allgemein gesagt Informationsmaterial eindeutig, basierend auf dem Inhalt, zu identifizieren.

[0003] Eine Anwendung hierfür ist das sogenannte „Broadcast-Monitoring“ bzw. die Rundfunküberwachung. Mit Hilfe eines solchen Audio-Video-Überwachungssystems soll z. B. sichergestellt werden, daß nur rechtmäßige Inhalte verbreitet werden oder daß die jeweiligen Tantiemen für die Rechte-Inhaber des Audio- und Video-Materials korrekt abgerechnet werden.

[0004] Eine weitere Anwendung ist z. B. die Erkennung von Audiomaterial, welches über Peer-to-Peer-Netzwerke zwischen Partnern ausgetauscht werden soll.

[0005] Eine weitere Anwendung ist die Kontrollmöglichkeit der Werbeindustrie, um einen Fernseh- oder Radiosender dahingehend zu kontrollieren, ob die gebuchten Werbezeiten auch tatsächlich ausgestrahlt worden sind, oder ob nur Teile der gebuchten Werbeanteile ausgesendet worden sind, oder ob Teile der Werbesendungen während der Übertragung gestört worden sind, was z. B. in die Verantwortung des Fernseh- oder Rundfunk-Senders fallen könnte. An dieser Stelle sei darauf hingewiesen, daß insbesondere die Kosten für die Fernsehwerbung in weit verbreiteten Programmen zu guten Sendezeiten derart hoch sind, daß die Werbeindustrie insbesondere im Hinblick auf diese hohen Kosten ein vitales Interesse an einer Kontrollmöglichkeit hat, um nicht nur auf die Angaben der Rundfunksender vertrauen zu müssen. Bisher wird als Kontrollmöglichkeit auf bezahlte „Probehörer“ bzw. „Probehörerinnen“ oder „Probeseher“ bzw. „Probeseherinnen“ gesetzt, die ein bestimmtes Fernsehprogramm durchgehend betrachten und z. B. die genauen Zeiten protokollieren, zu denen ein Werbespot übertragen wird und die ferner überwachen, ob während der Übertragung keine Störung aufgetreten ist, oder ob der gesamte Werbespot korrekt übertragen worden ist, ob also keine Bildverzerrung stattgefunden hat, etc.

[0006] Die Nachteile dieses Konzepts sind evident.

So sind zum einen die Kosten erheblich, und ist zum anderen die Verlässlichkeit oder Beweiskraft von Aussagen der Probehörer bzw. Probeseher problematisch, insbesondere wenn Rückzahlungsforderungen in beträchtlicher Höhe erhoben werden, welche im Hinblick auf ihre Nachweisbarkeit allein an Probebeobachtern bzw. Probebeobachterinnen hängt.

[0007] Zum automatisierten Broadcast-Monitoring können verschiedene bekannte Systeme eingesetzt werden. So offenbart die WO 02/11123 A2 oder die Fachveröffentlichung: „Invited Talk: An Industrial-Strength Audio Search Algorithm“, Avery Wang, ISMIR 2003, Baltimore, Oktober 2003, Systeme und Verfahren zum Erkennen von Ton- und Musiksignalen in einer Umgebung starken Rauschens und hoher Verzerrungen. Hierbei wird zunächst untersucht, ob eine Übereinstimmung zwischen Hash-Werten eines Referenz-Audio-Objekts und dem aktuell ermittelten Hash-Wert des noch unidentifizierten Audio-Objekts vorliegt. Ist dies der Fall, so wird der zugehörige Zeitoffset, also der relative Abstand vom Beginn des Audio-Objekts, des Hash-Werts in dem noch unidentifizierten Audio-Objekt und der Zeitoffset des Hash-Werts in dem Referenz-Audio-Objekt unter der jeweiligen Kennung des Referenz-Audio-Objekts gespeichert. Sind alle Eingangshashwerte verarbeitet, dann beginnt eine sogenannten Scanning-Phase. In dieser wird untersucht, wie viele Zeitoffsetpaare pro Referenz-Audio-Objektzeit kontinuierlich zusammenpassen. Wird eine bestimmte Anzahl festgestellt, so geht man von einer Identifizierung des entsprechenden Referenz-Audio-Objekts aus. Die Zeitoffsetpaare werden genau dann als zeitkontinuierlich, d. h. zeitlich zueinander gehörig, angesehen, wenn sie in einem zweidimensionalen Scatterplot mit einem Zeitoffset als x- und dem anderen als y-Koordinate eine Gerade bilden.

[0008] In der Fachveröffentlichung „Robust Audio Hashing for Content Identification“ von J. Haitzma, T. Kalker, J. Oostveen, in den Proceedings of the Content-Based Multimedia Indexing, 2001, url:cite-seer.ist.psu.edu/haitzma01robust.html, wird ein System zum robusten Audio-Hashing für eine Inhalt-identifikation dargestellt. Zur inhaltsbasierten Musikkennung wird eine Hash-Funktion eingesetzt, welche eine Bitfolge mit einem Abschnitt aus einem Audiosignal assoziiert, und zwar in einer Form, daß für die menschliche Schallwahrnehmung akustisch ähnliche Audiosignale auch eine ähnliche Bitfolge erzeugen. Zur Berechnung eines Hash-Werts wird zunächst das Audiosignal gefenstert und einer Transformation unterzogen, um schließlich eine Einteilung des Transformationsergebnisses in Frequenzbänder mit logarithmischer Bandbreite vorzunehmen. Für diese Frequenzbänder werden die Vorzeichen der Differenzen in Zeit- und Frequenzrichtung bestimmt. Die sich aus den Vorzeichen ergebende Bitfolge bildet den Hash-Wert. Es wird immer ein Hash-Wert für

eine Audiosignallänge von 3 Sekunden berechnet. Liegt für einen solchen Abschnitt die Hamming-Distanz zwischen einem Referenz-Hash-Wert und einem zu untersuchenden Test-Hash-Wert unterhalb einer Schwelle s , so wird eine Übereinstimmung angenommen und der Testabschnitt dem Referenzelement zugeordnet.

[0009] Um eine Erkennung von Audiomaterial durchzuführen, wird das Audiosignal typischerweise in kleine Einheiten der Länge Δt zerlegt. Diese einzelnen Einheiten werden jeweils einzeln analysiert, um wenigstens eine gewisse zeitliche Auflösung zu haben.

[0010] Hieraus ergeben sich mehrere Probleme.

[0011] Die Erkennungsergebnisse der kleinen analysierten Zeitabschnitte des Audiosignals müssen so zusammengesetzt werden, daß für einen längeren Zeitabschnitt eine eindeutige, korrekte Aussage über das erkannte Audiosignal getroffen werden kann.

[0012] Zur Analyse eines kontinuierlichen Audiodatenstroms sollen Übergänge von einem Audioelement zu einem anderen richtig erkannt werden, d. h. ein Wechsel von einem Musikstück A zu einem Musikstück B.

[0013] Ferner existiert die Situation, bei der von einem Musikstück mehrere Versionen vorkommen, die z. B. gleich beginnen und sich erst nach einer gewissen Zeit anfangen zu unterscheiden. Hierbei sei beispielsweise an Kurzversionen oder Maxiversionen eines Songs gedacht. Alternativ existieren auch Situationen dahingehend, daß Musikstücke, die auf den gleichen Song zurückgehen, z. B. am Anfang unterschiedlich sind, einen identischen Mittelteil haben und sich gegen Ende wenigstens eines der beiden Musikstücke voneinander wieder unterscheiden. Für die Tantiemenvergabe an Urheberrecht-Inhaber kann es durchaus eine Rolle spielen, ob z. B. gegen eine höhere Gebühr die Maxiversion eines Songs gespielt werden kann, ob für eine mittlere Gebühr nur eine Normalversion gespielt werden darf, oder ob für eine geringe Gebühr bereits die Kurzversion eines Songs gespielt werden kann. In diesem Fall sollten verschiedene Versionen eines Songs zuverlässig voneinander unterschieden werden können.

[0014] Der vorstehende Stand der Technik ist dahingehend unzureichend, daß sich Erfassungsfehler ergeben, wenn die Ergebnisse der Einzelerkennungen einfach zusammengesetzt werden. Insbesondere werden keine Angaben dahingehend geliefert, ob und wie ein kontinuierlicher Audiodatenstrom aus mehreren verschiedenen Audioobjekten analysiert werden kann und wie entsprechende Wechsel zwischen verschiedenen Audioobjekten detektiert werden können. Des weiteren wird insbesondere im

letztgenannten Stand der Technik die Vieldeutigkeit von Referenz-Hash-Werten zwar erwähnt. Es wird jedoch keine explizite Lösung für das Problem zur Bestimmung eines eindeutigen Kandidaten genannt. Gilt für einen Hash-Wert ein Audioobjekt als identifiziert, so wird für den unmittelbar folgenden Hash-Wert lediglich geprüft, ob dieser zu dem identifizierten Audioobjekt paßt. Ist dies nicht der Fall, so wird neu gesucht, unter Einbeziehung von allen Referenzaudioobjekten.

[0015] Insbesondere zur Unterscheidung von unterschiedlichen Versionen ein und desselben Songs ist im Stand der Technik keine Lösung bekannt.

[0016] Die WO 01/04870 A1 offenbart ein Verfahren zur automatischen Erkennung von Musikkompositionen und Tonsignalen. Ein unbekanntes Musiksignal wird gefenstert und einer Merkmalsextraktionsprozedur unterzogen. Die extrahierten Merkmale werden mit Merkmalen verglichen, die von einem bekannten Stück abgeleitet worden sind. Insbesondere wird für aufeinander folgende gefensterte Blöcke des unbekanntes Musiksignals das Spektrum berechnet, um dann vom Betrag des Spektrums die Spitzen zu erhalten. Die Spitzen werden in vorgegebene Frequenz-Bins eingeordnet, und die Ordnungszahlen der Frequenz-Bins, die die L-größten Spitzen haben, dienen als Repräsentationsvektor. Der Repräsentationsvektor wird mit einem entsprechenden Repräsentationsvektor eines bekannten Stücks verglichen und dann, wenn mehr als die Hälfte der Vektoreinträge übereinstimmen, einer weiteren Prüfung unterzogen. Stimmen jedoch weniger als die Hälfte der Vektorelemente überein, so wird die Prüfung des Musikstücks, von dem der Vektor stammt, abgebrochen und mit einem neuen Vektor eines anderen bekannten Musikstücks begonnen.

Aufgabenstellung

[0017] Die Aufgabe der vorliegenden Erfindung besteht darin, ein zuverlässiges Konzept zum Analysieren eines Informationssignals zu schaffen.

[0018] Diese Aufgabe wird durch eine Vorrichtung zum Analysieren eines Informationssignals gemäß Patentanspruch 1, ein Verfahren zum Analysieren eines Informationssignals gemäß Patentanspruch 20 oder ein Computer-Programm gemäß Patentanspruch 21 gelöst.

[0019] Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, daß eine zuverlässige Inhaltsidentifikation dadurch erreicht wird, daß nicht nur Einzelerkennungsergebnisse für sich betrachtet berücksichtigt werden, sondern über einen gewissen Zeitraum hinweg. So existiert auch in der Folge von Einzelerkennungsergebnissen für eine Folge von Fingerabdrücken eine erhebliche Information, die zur Erkennung

einsetzbar ist. Erfindungsgemäß wird daher aus einer Folge von Fingerabdrücken, die eine Folge von Blöcken eines Informationssignals darstellt, eine Bildung von wenigstens zwei unterschiedlichen Hypothesen durchgeführt, wobei eine erste Hypothese eine Annahme für die Zugehörigkeit der Folge von Blöcken zu einer ersten Informationsentität ist, und wobei die zweite Hypothese eine Annahme für die Zugehörigkeit der Folge von Blöcken zu der zweiten Informationsentität ist. Die wenigstens zwei Hypothesen werden nunmehr untersucht und einer Bewertung dahingehend unterzogen, daß eine Aussage über das Informationssignal basierend auf einem Untersuchungsergebnis getroffen wird. Die Aussage könnte beispielsweise darin bestehen, daß bestimmt wird, daß die Folge von Blöcken eine Informationsentität darstellt, die eine Hypothese hat, die am wahrscheinlichsten ist. Die Aussage könnte alternativ oder zusätzlich darin bestehen, daß eine Informationseinheit mit dem Fingerabdruck beendet ist, der als zeitlich letzter Fingerabdruck aus der Folge von Fingerabdrücken zu der wahrscheinlichsten Hypothese beiträgt.

[0020] Vorzugsweise werden die Hypothesen dahingehend untersucht, daß es für Fingerabdrücke wenigstens zwei unterschiedliche Identifikationsergebnisse gibt, sowie daß es für jeden der zwei unterschiedlichen Identifikationsergebnisse ein Zuverlässigkeitsmaß gibt, wobei dieses Zuverlässigkeitsmaß in einer konkreten Zahl bestehen kann. Dieses Zuverlässigkeitsmaß kann jedoch auch implizit dahingehend gegeben sein, daß allein die Tatsache, daß z. B. zwei Identifikationsergebnisse geliefert werden, eine Zuverlässigkeit von z. B. 1/2 signalisiert wird und daß diese Zahl nicht explizit angegeben wird.

[0021] Zur Beurteilung, ob eine Hypothese wahrscheinlicher ist als die andere Hypothese werden vorzugsweise Zuverlässigkeitsmaße der Einzelerkennungen für die betreffende Anzahl von zeitlich aufeinanderfolgenden Blöcken zusammengefaßt, wobei diese Zusammenfassung vorzugsweise in einer Aufaddition bestehen wird. Hierauf wird die Hypothese, die das höchste zusammengefaßte Zuverlässigkeitsmaß liefert, als die wahrscheinlichste Hypothese bewertet.

[0022] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird als Mittel zum Bereitstellen von aufeinanderfolgenden Identifikationsergebnissen eine Fingerabdruck-Datenbank verwendet, in der eine Anzahl von Referenz-Fingerabdrücken jeweils in Zuordnung zu einem Identifikationsergebnis abgelegt sind. Hierauf wird mit dem aus einem Block des zu analysierenden Informationssignals erzeugten Fingerabdrucks eine Datenbanksuche vorgenommen, um innerhalb der Datenbank nach einem Referenzfingerabdruck zu suchen, der eine Übereinstimmung mit dem Test-Fingerabdruck liefert. Je

nach Ausgestaltung der Datenbank wird als Suchergebnis nur der beste Treffer, also der Treffer mit einem minimalen Abstandsmaß von der Datenbank als Identifikationsergebnis ausgegeben. Es werden auch Datenbanken bevorzugt, die nicht nur qualitativ ein Trefferergebnis liefern, sondern auch ein quantitatives Trefferergebnis liefern, dahingehend, daß eine Anzahl möglicher Treffer mit einem zugeordneten Zuverlässigkeitsmaß ausgegeben wird, so daß beispielsweise alle Treffer mit einem Zuverlässigkeitsmaß, das höher oder gleich einer bestimmten Schwelle, wie beispielsweise 20% ist, von der Datenbank ausgegeben werden.

[0023] Bei dem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird dann eine neue Hypothese begonnen, wenn ein neues Identifikationsergebnis auftaucht, für das noch keine Hypothese existiert. Dieses Prozedere wird für eine bestimmte Anzahl von Blöcken durchgeführt, um dann in die Vergangenheit gerichtet zu untersuchen, ob eine bestimmte sich als zuverlässig herausgestellte Hypothese bereits zu Ende ist, um dann diese Hypothese als die wahrscheinlichste Hypothese zu erkennen.

[0024] Ein Vorteil der vorliegenden Erfindung besteht darin, daß das Konzept zuverlässig arbeitet und dennoch fehlertolerant insbesondere im Hinblick auf Übertragungsfehler ist. So wird nicht anhand eines einzigen Blockes versucht eine Entscheidung herbeizuführen, sondern es wird eine Folge von aufeinanderfolgenden Blöcken durch Hypothesenbildung gewissermaßen gemeinsam betrachtet und ausgewertet, so daß nicht kurzfristige Übertragungsstörungen bzw. allgemeines auftretendes Rauschen den gesamten Erkennungsprozeß hinfällig machen.

[0025] Weiterhin liefert das erfindungsgemäße Konzept automatisch eine Protokollierung der Übertragungsqualität vom Anfang bis zum Ende z. B. eines Werbespots. Auch wenn eine Hypothese als wahrscheinlichste Hypothese erkannt worden ist, wenn also bestimmt wird, daß ein bestimmter Werbespot da war, so sind dennoch Qualitätsschwankungen innerhalb des Werbespots anhand der Zuverlässigkeitsmaße nachvollziehbar. Ferner ist damit insbesondere die komplette zeitliche Kontinuität eines Werbespots als Beispiel für eine Informationsentität nachvollziehbar und protokollierbar, und zwar insbesondere dahingehend, daß nicht irgend etwas aus dem Werbespot ständig wiederholt worden ist, sondern daß der gesamte Werbespot vom Anfang des Werbespots bis zum Ende des Werbespots kontinuierlich gesendet worden ist.

[0026] Die vorliegende Erfindung ist ferner dahingehend vorteilhaft, daß durch Hypothesenbildung automatisch das Ende einer Informationsentität und der Beginn einer Informationsentität erkannt werden. Dies liegt daran, daß eine Zuordnung zu einer Infor-

mationsentität im allgemeinen eindeutig sein wird. Dies bedeutet, daß über einen bestimmten Zeitpunkt nicht mehrere Informationsentitäten gemeinsam abgespielt werden können, sondern daß zumindest für die übergroße Anzahl von Programminhalten nur eine Informationsentität im Informationssignal zu einem Zeitpunkt enthalten ist. Die Hypothesenuntersuchung und die Beurteilung der Hypothesen aufgrund der Hypothesenuntersuchung liefert automatisch einen Zeitpunkt, zu dem eine vorherige Informationsentität beendet ist und zu dem eine neue Informationsentität beginnt. Dies liegt an der bis in die Hypothesen hinein beibehaltenen Blockzuordnung. So entspricht nach wie vor eine Folge von Fingerabdrücken einer Folge von Blöcken, und entspricht wiederum eine Folge von Identifikationsergebnissen einer Folge von Fingerabdrücken, so daß eine Hypothese dem ursprünglichen Informationssignal in zeitlicher Hinsicht eindeutig zugeordnet ist.

[0027] Das erfindungsgemäße Konzept ist ferner dahingehend vorteilhaft, daß keine „Unentschieden“-Situationen zwischen zwei Hypothesen entstehen, auch wenn Informationsentitäten streckenweise dennoch das identische Audiomaterial haben, wie beispielsweise bei Kurzversionen oder Langversionen ein und desselben Songs.

Ausführungsbeispiel

[0028] Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend Bezug nehmend auf die beiliegenden Zeichnungen detailliert erläutert. Es zeigen:

[0029] [Fig. 1](#) ein Blockschaltbild einer erfindungsgemäßen Vorrichtung;

[0030] [Fig. 2](#) ein Blockschaltbild einer für das in [Fig. 1](#) gezeigte Ausführungsbeispiel verwendbaren Datenbank;

[0031] [Fig. 3](#) eine schematische Darstellung eines Ausgabeergebnisses für eine Folge von Fingerabdrücken für eine Folge von Zeitintervallen sowie die zugeordneten Hypothesen;

[0032] [Fig. 4a-Fig. 4c](#) ein Beispielszenario für nachfolgende Anwendungsbeispiele;

[0033] [Fig. 5a-Fig. 5d](#) eine schematische Darstellung verschiedener Fehlinterpretationen;

[0034] [Fig. 6](#) ein Blockschaltbild eines bevorzugten Ausführungsbeispiels der vorliegenden Erfindung;

[0035] [Fig. 7a-Fig. 7c](#) eine Darstellung der Funktionalität des erfindungsgemäßen Konzepts für das in den [Fig. 4a-Fig. 4c](#) dargestellte Ausgangsszenario;

[0036] [Fig. 8](#) eine schematische Darstellung eines Informationssignals mit Informationseinheiten, Blöcken von Informationseinheiten und Informationsentitäten mit einer Mehrzahl von Blöcken;

[0037] [Fig. 9](#) ein bekanntes Szenario zum Aufbauen einer Fingerabdruck-Datenbank; und

[0038] [Fig. 10](#) ein bekanntes Szenario zum Audio-Identifizieren mittels einer gemäß [Fig. 9](#) geladenen Fingerabdruck-Datenbank.

[0039] [Fig. 1](#) zeigt ein Blockschaltbild einer Vorrichtung zum Analysieren eines Informationssignals gemäß einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung. Ein beispielhaftes Informationssignal ist bei **800** in [Fig. 8](#) dargestellt. Das Informationssignal **800** besteht aus einer Folge **802** von zeitlich aufeinanderfolgenden Blöcken von Informationseinheiten, wobei die einzelnen Informationseinheiten **804** z. B. Audio-Abtastwerte, Video-Pixels oder Video-Transformationskoeffizienten etc. sein könnten. Eine Mehrzahl von Blöcken der Folge **802** bilden zusammen immer eine Informationsentität **806**. Bei dem in [Fig. 8](#) gezeigten Ausführungsbeispiel bilden die ersten sechs Blöcke die erste Informationsentität, und bilden die Blöcke **7, 8, 9, 10** die zweite Informationsentität. Ab den Blöcken **11** bis **n** ist in [Fig. 8](#) z. B. eine dritte Informationsentität dargestellt. Eine Informationsentität könnte beispielsweise ein Musikstück, eine gesprochene Passage, ein Videobild oder auch z. B. ein Teil eines Videobildes sein. Eine Informationsentität könnte jedoch auch ein Text oder z. B. eine Seite eines Textes sein, wenn das Informationssignal auch Textdaten umfaßt.

[0040] Die in [Fig. 1](#) gezeigte Vorrichtung ist ausgebildet, um unter Verwendung einer Folge von Fingerabdrücken FA1, FA2, FA3, ..., FAi zu arbeiten, die aus der Folge von Blöcken **802** erzeugt werden, oder die z. B. von einem Speicher abgerufen werden, wenn die Fingerabdrücke bereits vor der Analyse erzeugt worden sind oder vielleicht sogar mit dem Informationssignal mitgeliefert werden, je nach Ausführungsform. Es sei darauf hingewiesen, dass bei der Blockbildung auch Blocküberlappungstechniken eingesetzt werden können, wie sie z. B. aus der Audiocodierung bekannt sind.

[0041] Auf jeden Fall arbeitet die Vorrichtung zum Analysieren des Informationssignals unter Verwendung einer Folge von Fingerabdrücken für die Folge von Blöcken, so daß die Folge von Blöcken **802** durch die Folge von Fingerabdrücken FA1, FA2, FA3, FA4, ..., FAi dargestellt ist. Die Folge von Fingerabdrücken wird in einen Fingerabdruck-Eingang in eine Einrichtung **12** zum Bereitstellen von Identifikationsergebnissen für aufeinanderfolgende Fingerabdrücke eingespeist. Die Einrichtung **12** zum Bereitstellen von aufeinanderfolgenden Identifikationsergebnis-

sen ist wirksam, um für die aufeinanderfolgenden Fingerabdrücke aufeinanderfolgende Identifikationsergebnisse zu liefern, wobei ein Identifikationsergebnis eine Zugehörigkeit eines Blocks von Informationseinheiten zu einer vorbestimmten Informationsentität darstellt. Wenn z. B. davon ausgegangen wird, daß ein Song eine zeitliche Länge hat, die in etwa sechs Blöcken entspricht, so liefern die sechs Blöcke zwar unterschiedliche Fingerabdrücke, in der Einrichtung **12** zum Bereitstellen ist jedoch signalisiert, daß alle diese sechs Blöcke zu der vorbestimmten Informationsentität, also zu dem besagten Lied gehören.

[0042] Je nach Ausführungsform wird die Einrichtung **12** zum Bereitstellen für einen Fingerabdruck einen oder mehrere Identifikationsergebnisse liefern. Dieser eine oder die mehreren Identifikationsergebnisse werden einer Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen aus den Identifikationsergebnissen für die aufeinanderfolgenden Fingerabdrücke zugeführt. Insbesondere stellt eine erste Hypothese eine Annahme für die Zugehörigkeit der Folge von Blöcken zu einer ersten Informationsentität dar, und ist die zweite Hypothese eine Annahme für die Zugehörigkeit der Folge von Blöcken zu der zweiten Informationsentität. Die verschiedenen Hypothesen H1, H2,..., werden einer Einrichtung **16** zum Untersuchen der Hypothesen zugeführt, wobei die Einrichtung **16** ausgebildet ist, um gemäß eines einstellbaren Untersuchungsalgorithmus zu arbeiten, um schließlich ein Untersuchungsergebnis an einem Untersuchungsergebnis Ausgang **18** zu liefern.

[0043] Dieses Untersuchungsergebnis an der Leitung **18** wird dann einer Einrichtung **20** zum Treffen einer Aussage über das Informationssignal zugeführt. Die Einrichtung **20** zum Treffen einer Aussage über das Informationssignal ist ausgebildet, um eine Information über das Informationssignal basierend auf dem Untersuchungsergebnis auszugeben und kann vielerlei Einstellungen haben.

[0044] Für alle Einstellungen ist gemeinsam, daß die Aussage über das Informationssignal auf der Basis des Untersuchungsergebnisses **18** getroffen wird. Beispiele für verschiedene Aussagen über das Informationssignal bestehen darin, daß bestimmt wird, daß die Folge von Blöcken eine Informationsentität darstellt, die eine Hypothese hat, die am wahrscheinlichsten ist. Alternative Aussagen bestehen darin, daß eine Informationsentität mit dem Fingerabdruck beendet ist, der als zeitlich letzter Fingerabdruck zu der wahrscheinlichsten Hypothese beiträgt. Eine alternative Aussage, die durch die Einrichtung **20** getroffen werden kann, besteht darin, daß festgestellt wird, daß in dem Informationssignal eine Informationsentität an sich vorhanden ist oder nicht.

[0045] Das erfindungsgemäße Postprocessing, das insbesondere durch die Einrichtungen **14**, **16** und **20**

geliefert wird, also das Bilden von wenigstens zwei Hypothesen, das Untersuchen der Hypothesen und das Treffen einer Aussage auf der Basis eines Untersuchungsergebnisses ermöglicht somit nicht nur die Identifikation eines Stücks in einem an sich unbekanntem, also zu analysierendem Informationssignal, sondern ermöglicht auch – abgesehen von der Identifikation eines Stücks an sich – die Erkennung des Endes eines ersten Stücks, also einer ersten Informationsentität, und das Detektieren des Beginns einer zweiten, auf die erste Informationsentität folgenden Informationsentität.

[0046] Im Hinblick auf die Werbeüberwachung liefert das erfindungsgemäße Postprocessing-Konzept jedoch auch die Möglichkeit, zu erfassen, ob ein bestimmtes Stück in dem Informationssignal vorhanden war oder nicht. Hierbei würden die aus dem Informationssignal gewonnenen Fingerabdrücke lediglich mit einem Satz von Fingerabdrücken, nämlich dem Satz von Fingerabdrücken, die die vorbestimmte Informationsentität, also einen bestimmten Werbespot darstellen, verglichen. Diese Aussage ist somit nicht primär dahingehend zu sehen, daß eine Informationsentität identifiziert wird, oder daß das Ende einer Informationsentität und der Beginn einer folgenden Informationsentität erfaßt werden, sondern besteht darin, zu erfassen, ob eine bestimmte Informationsentität in einem an sich unbekanntem zu analysierendem Informationssignal vorhanden ist oder nicht.

[0047] [Fig. 2](#) zeigt eine spezielle bevorzugte Ausführungsform der Einrichtung **12** zum Bereitstellen von Identifikationsergebnissen für aufeinanderfolgende Fingerabdrücke. Bei einem bevorzugten Ausführungsbeispiel umfaßt die Einrichtung **12**, wie es in [Fig. 2](#) gezeigt ist, eine Datenbank, die verschiedene Referenzfingerabdrücke FARj umfaßt, die alle in Zuordnung zu einem Identifikationsergebnis, d. h. IDk, abgespeichert sind. Bei dem bevorzugten Ausführungsbeispiel werden die Fingerabdrücke FAi nach und nach, also zeitlich sequentiell verarbeitet. So wird ein Fingerabdruck FAi in die Datenbank über eine Eingangsleitung **24** eingespeichert. In der Datenbank wird dann der eingespeicherte Fingerabdruck FAi mit allen Referenzfingerabdrücken FARj verglichen. bei dem bevorzugten Ausführungsbeispiel ist die Datenbank nicht eine qualitative Datenbank, die feststellt, daß ein Eingangsfingerabdruck mit einem gespeicherten Referenzfingerabdruck übereinstimmt oder nicht, sondern ist die Datenbank eine quantitative Datenbank, die ein Abstandsmaß bzw. ein Zuverlässigkeitsmaß für die ausgegebenen Ergebnisse liefern kann. So würde die Datenbank **22** bei dem in [Fig. 2](#) gezeigten bevorzugten Ausführungsbeispiel an ihrem Ausgang **26** z. B. das Ergebnis liefern, das in einer Ergebnistabelle **28** dargestellt ist. So würde die Datenbank z. B. sagen, daß der Fingerabdruck FAi auf ein Identifikationsergebnis IDx, also auf ein Musikstück beispielsweise x mit einer Zu-

verlässigkeit ZV_1 von 60% hinweist. Gleichzeitig wird die Datenbank jedoch auch sagen, daß der Fingerabdruck FAi mit einer Zuverlässigkeit von 50% auf ein Stück mit dem Identifikationsergebnis IDy hinweist. Schließlich könnte die Datenbank auch ausgeben, daß der Fingerabdruck FAi mit einem Zuverlässigkeitsmaß ZV_3 von beispielsweise 40% auf ein wieder anderes Stück mit der Identifikation IDz hinweist.

[0048] Je nach Ausführungsform kann die gesamte Ergebnistabelle **28** der Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen von [Fig. 1](#) zugeführt werden. Alternativ könnte jedoch die Datenbank **22** selbst bereits eine Entscheidung treffen und immer nur den wahrscheinlichsten Wert, also im vorliegenden Fall das Ergebnis IDx der Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen zuführen. In diesem Fall müßte nicht unbedingt das Zuverlässigkeitsmaß ZV_1 ebenfalls der Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen zugeführt werden. Vielmehr könnte auf die weitere Übermittlung der Zuverlässigkeitsmaße ZV_i verzichtet werden. Alternativ könnte jedoch auch die Einrichtung **12** zum Bereitstellen der Identifikationsergebnisse, die gleichzeitig auch die Zuverlässigkeitsmaße liefert, ausgebildet sein, um die Zuverlässigkeitsmaße ZV_i in entsprechender Reihenfolge in Zuordnung zu den Blöcken nicht der Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen zuzuführen, sondern der Einrichtung **16** zum Untersuchen der Hypothesen, da diese Einrichtung **16** die Zuverlässigkeitsmaße erst benötigt, um z. B. die wahrscheinlichste Hypothese zu finden.

[0049] Aus der Datenbank **22** in [Fig. 2](#) ist zu sehen, daß ein Identifikationsergebnis, wie z. B. ID1 mehrere zugeordnete Fingerabdrücke FAR11, FAR12, FAR13 haben kann, was darauf hindeutet, daß das Stück, das durch ID1 identifiziert wird, mehrere Blöcke hat. Je nach Implementierung kann jedoch auch ein einziger langer Fingerabdruck für das Stück mit der Identifikation ID1 gespeichert sein, der sich jedoch aus den einzelnen Fingerabdrücken FAR11, FAR12, FAR13,... zusammensetzt. Die Datenbank würde dann den eing gespeisten Fingerabdruck FAi, der von der Blocklänge abhängt und typischerweise viel kürzer ist als der lange Fingerabdruck, mit dem langen Fingerabdruck in jeder Zeile der Datenbank korrelieren, um festzustellen, ob ein Ausschnitt des langen abgespeicherten Referenzfingerabdrucks mit dem an der Leitung **24** eingespeisten Referenzfingerabdruck FAi übereinstimmt oder nicht. Das Zuverlässigkeitsmaß würde sich hierbei gewissermaßen automatisch ergeben, nämlich einfach durch quantitative Auswertung des Korrelationsergebnisses.

[0050] Ferner sei Bezug nehmend auf [Fig. 2](#) bereits auf die beiden letzten Zeilen verwiesen, die mit den Identifikationsergebnissen ID108 und ID109 bezeichnet sind. ID108 bezeichnet eine lange Version eines

Musikstücks, wie es noch Bezug nehmend auf [Fig. 4a](#) erläutert wird, während ID109 eine kurze Version desselben Musikstücks identifiziert, wie es in [Fig. 4b](#) dargestellt ist.

[0051] Wie es bereits ausgeführt worden ist, kann die Datenbank **22**, also diese Implementierung der Einrichtung **12** zum Bereitstellen von Identifikationsergebnissen für aufeinanderfolgende Fingerabdrücke derart ausgebildet sein, daß sie immer nur das wahrscheinlichste Identifikationsergebnis liefert. Alternativ könnte die Datenbank **22** jedoch auch ausgebildet sein, um z. B. immer nur die Identifikationsergebnisse zu liefern, deren Wahrscheinlichkeit höher als eine Mindestschwelle ist, wie beispielsweise eine Schwelle von 5%. Dies würde dazu führen, daß die Anzahl der Zeilen der Tabelle von Fingerabdruck zu Fingerabdruck variieren wird. Wieder alternativ könnte die Datenbank **22** jedoch auch implementiert sein, um für jeden Eingangsfingerabdruck FAi eine bestimmte Anzahl von wahrscheinlichsten Kandidaten, wie beispielsweise die „Top Ten“, also die zehn wahrscheinlichsten Kandidaten der Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen zuzuführen.

[0052] Nachfolgend wird anhand von [Fig. 3](#) eine Ausführungsform der Datenbank **22** dargestellt, bei der die Datenbank immer die drei wahrscheinlichsten Identifikationsergebnisse samt zugehöriger Zuverlässigkeitswerte der Einrichtung **14** zum Bilden von Hypothesen zuführt, also gewissermaßen eine „Top Three“-Implementierung umfaßt. So ist aus [Fig. 3](#) zu sehen, daß für den Fingerabdruck FA1 Identifikationsergebnisse ID1, ID2, ID3 geliefert werden, und zwar mit den jeweiligen Zuverlässigkeitsmaßen 40%, 60% oder 30%. Für das Zeitintervall Δt_2 , also für den Fingerabdruck FA2 wird wieder eine Lieferung der Identifikationsergebnisse ID1, ID2, ID3 stattfinden, nun jedoch mit einer anderen jeweiligen Wahrscheinlichkeit, also mit einem anderen jeweiligen Zuverlässigkeitsmaß, das lediglich beispielhaft in [Fig. 3](#) prozentual dargestellt ist. Dieses Prozedere wird für sämtliche Eingangs-Fingerabdrücke FA1 bis FA8 durchgeführt. Der Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen, wie es in [Fig. 1](#) dargestellt ist, werden diese Identifikationsergebnisse geliefert. Die Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen ist ausgebildet, um immer dann, wenn ein neues Identifikationsergebnis von der Einrichtung **12** zum Bereitstellen der Identifikationsergebnisse geliefert wird, eine neue Hypothese zu beginnen. Dies ist aus

[0053] [Fig. 3](#) daraus ersichtlich, daß beim Zeitpunkt Δt_1 die Hypothesen H1, H2, H3 mit ID1, ID2 bzw. ID3 begonnen werden, und daß im Zeitintervall Δt_7 wieder neue Hypothesen mit ID108, ID109, ID4 begonnen werden, und daß im Zeitintervall Δt_8 aufgrund der Tatsache, daß dort bei dem gezeigten Beispiel ID8 zum ersten Mal auftritt, eine weitere Hypothese

H4 für ID8 begonnen wird.

[0054] Die Einrichtung **14** zum Bilden von wenigstens zwei Hypothesen ist daher wirksam, um für jeden neuen Fingerabdruck zu sehen, ob ein neues Identifikationsergebnis kommt, um eine neue Hypothese zu beginnen, und um eine bereits früher begonnene Hypothese dahingehend weiterzuführen, wenn für einen Zeitabschnitt Δt_i für die bereits früher begonnene Hypothese in den „Top Three“ oder „Top x“ ein Element enthalten ist, das, wenn auch mit geringerer Wahrscheinlichkeit dennoch ein Identifikationsergebnis für eine gerade begonnene Hypothese liefert. Dieses Prozedere wird eine bestimmte Zeit lang fortgeführt. Dann, beispielsweise zu vorbestimmten Zeitpunkten oder, getriggert durch einen Benutzer, etc. wird die Einrichtung **16** zum Untersuchen der Hypothesen die für die Vergangenheit gebildeten Hypothesen untersuchen und für den in [Fig. 3](#) gezeigten Fall beispielsweise die Zuverlässigkeitsmaße der Hypothesen H1, H2, H3 für die Zeiträume Δt_1 bis Δt_6 addieren. Die Einrichtung **16** zum Untersuchen von wenigstens zwei Hypothesen würde dann feststellen, daß das Stück am wahrscheinlichsten ID1 ist, daß also die Hypothese H1 für den Zeitraum Δt_1 bis Δt_6 die wahrscheinlichste Hypothese ist, da das Zuverlässigkeitsmaß einen Wert von 420 erreicht, während für die zweite Hypothese lediglich ein Zuverlässigkeitsmaß von 230 erreicht wird, und während für die dritte Hypothese lediglich ein Zuverlässigkeitsmaß von 135 erreicht wird.

[0055] Bei dem in [Fig. 3](#) gezeigten Fall starten alle drei Hypothesen gleichzeitig und enden alle drei Hypothesen gleichzeitig. Dies muß jedoch nicht zwingend so sein. So könnte die Hypothese H1 z. B. früher enden, also z. B. zum Zeitpunkt Δt_5 . Das Zuverlässigkeitsmaß von ID1 müßte in diesem Fall um 90 reduziert werden, womit man bei einem Wert von 330 ankommen würde. In diesem Fall würde sich ergeben, daß die Hypothese H1 trotzdem die wahrscheinlichste Hypothese ist, obgleich Hypothese H2 über einen längeren Zeitraum, jedoch insgesamt mit geringerer Wahrscheinlichkeit vorhanden ist. Aus dem in [Fig. 3](#) gezeigten Beispiel ist ferner zu sehen, daß die Hypothese H1 trotz der Tatsache, daß sie für Δt_1 weniger wahrscheinlich war als die Hypothese H2, am Ende „die Nase vorn“ hat.

[0056] Aus [Fig. 3](#) ist ferner auch zu sehen, daß eine Hypothese auch „Löcher“ dahingehend haben könnte, daß beispielsweise im Zeitintervall Δt_4 aus irgendwelchen Gründen, beispielsweise aus Gründen der Störung eines Übertragungskanal etc. lediglich ID2 und ID3 mit vernünftiger Wahrscheinlichkeit geliefert werden, nicht jedoch ID1. Dann müßte der Zuverlässigkeitswert für ID1 um 60 reduziert werden, was wiederum dahingehend zum Ergebnis haben würde, daß die Gesamtzuverlässigkeit statt 420 360 wäre, so daß auch in diesem Fall die Hypothese H1 die wahr-

scheinlichste Hypothese ist.

[0057] Aus den vorstehend beschriebenen Szenarien ist daher zu sehen, daß das erfindungsgemäße Konzept, das auf der Basis eines Postprocessings mit Hypothesen arbeitet und einerseits die Sequenz und andererseits die Zuverlässigkeitsmaße der einzelnen Fingerabdruckidentifikationsvorgänge berücksichtigt, außerordentlich robust gegenüber Übertragungsfehlern und auch gegenüber problematischen Funktionalitäten in der Datenbank oder auch gegenüber Fingerabdrücken ist, die sich für manche Informationsentitäten, wie z. B. Musikstücke, Videobilder, Texte, etc. vielleicht nicht so stark voneinander unterscheiden, wie man es sich wünschen würde.

[0058] Bei einem bevorzugten Ausführungsbeispiel ist eine Hypothese ein gespeichertes Protokoll ([Fig. 3](#): H1, H2, H3,...), vorzugsweise in der Form einer gespeicherten Liste, das einerseits einen Hinweis auf die Informationsentität, für die die Hypothese aufgestellt wird, und andererseits einen Hinweis auf Fingerabdrücke bzw. Blöcke von Informationseinheiten, für die die Hypothese geführt wird, aufweist. Vorzugsweise enthält das Protokoll auch noch für einen Block bzw. Fingerabdruck ein Zuverlässigkeitsmaß.

[0059] Aus [Fig. 3](#) ist ferner zu erkennen, daß sich die erste Informationsentität nur über den Zeitraum Δt_1 bis Δt_6 erstreckt und ab Δt_7 eine neue Entität beginnt. Dies ist insbesondere auch daraus ersichtlich, daß alle drei Hypothesen gleichzeitig enden, bzw. daß, auch wenn die Hypothese H3 beispielsweise sich noch über Δt_7 erstreckt hätte, nunmehr völlig andere Identifikationswerte mit sehr hoher Wahrscheinlichkeit, nämlich ID108 und ID109 mit Wahrscheinlichkeiten von 90 und 85 auftreten und damit die „klaren Gewinner“ aus dem vorherigen Zeitraum „ablösen“.

[0060] Am Ende von [Fig. 3](#) sind die verschiedenen Aussagen, die beispielhaft getroffen werden können, dargestellt, nämlich daß die Informationsentität im Zeitraum Δt_1 bis Δt_6 das durch ID1 identifizierte Musikstück ist. Alternativ könnte die Aussage auch dahingehend lauten, daß ein Informationsentität-Wechsel zwischen Δt_6 und Δt_7 stattfindet. Alternativ könnte eine Aussage jedoch auch dahingehend lauten, daß das Musikstück, das durch ID1 identifiziert ist, im Informationssignal enthalten ist.

[0061] Nachfolgend wird Bezug nehmend auf die [Fig. 9](#) und [Fig. 10](#) zunächst allgemeiner auf Datenbanksysteme eingegangen, wie sie in Verbindung mit der vorliegenden Erfindung vorteilhaft eingesetzt werden können. Die vorliegende Erfindung basiert somit auf einem System zu Identifikation von Audiomaterial, wie beispielsweise Musik. Das System kennt zwei Betriebsphasen. In der Trainings-Phase, die anhand von [Fig. 9](#) dargestellt ist, lernt das Erken-

nungssystem die später zu identifizierenden Stücke. In der Identifikations-Phase, die in [Fig. 10](#) dargestellt ist, können die vorher trainierten Audiostücke wiedererkannt werden.

[0062] Um ein Stück Musik – oder auch jedes andere Tonsignal – zu identifizieren, wird aus diesem ein kompakter und einzigartiger Datensatz extrahiert, der auch als Fingerprint oder Signatur bezeichnet wird. Diese Extraktion findet in einer Block Merkmalsextraktion **900** statt. In der Trainings- oder Einlernphase werden von einem Satz bekannter Audioobjekte solche Fingerprints erstellt und in einer Fingerprint-Datenbank **902** abgelegt. Vorzugsweise ist die Merkmalsextraktionseinrichtung **900** ausgebildet, um als Merkmal das SFM-Merkmal zu verwenden, wobei SFM für „Spectral Flatness Measure“ steht. Selbstverständlich können auch andere Fingerabdruckzeugungssysteme bzw. Merkmalsextraktionsergebnisse eingesetzt werden. Es hat sich jedoch herausgestellt, daß Tonalitäts-bezogene Merkmale und insbesondere das SFM-Merkmal einerseits eine besonders gute Unterscheidungskraft und andererseits eine besonders gute Kompaktheit haben. Zu diesem Zweck wird jeder Block zunächst einer Zeit-Frequenz-Umsetzung unterzogen, um dann mit den aus der Zeit-Frequenz-Umsetzung erzeugten Werten gemäß folgender Gleichung ein SFM für einen Block zu berechnen.

$$SFM = \frac{\left[\prod_{n=0}^{N-1} X(n) \right]^{\frac{1}{N}}}{\frac{1}{N} \sum_{n=0}^{N-1} X(n)}$$

[0063] In dieser Gleichung stellt $X(n)$ das Quadrat eines Absolutwerts einer Spektralkomponente mit den Index n dar, wobei N die Gesamtanzahl der Spektralkoeffizienten eines Spektrums darstellt. Es ist aus der Gleichung zu sehen, daß das SFM-Maß gleich dem Quotienten aus dem geometrischen Mittelwert der Spektralkomponenten und dem arithmetischen Mittelwert der Spektralkomponenten ist. Es ist bekannt, daß der geometrische Mittelwert immer kleiner oder höchstens gleich dem arithmetischen Mittelwert ist, so daß das SFM einen Wertebereich zwischen 0 und 1 hat. In diesem Kontext zeigt ein Wert nahe 0 ein tonales Signal an, und zeigt ein Wert nahe 1 ein ziemlich rauschartiges Signal mit einer flachen Spektralkurve an. Es sei darauf hingewiesen, daß der arithmetische Mittelwert und der geometrische Mittelwert nur gleich sind, wenn alle $X(n)$ identisch sind, was einem vollständig atonalen, d. h. rauschartigen oder pulsartigen Signal entspricht. Wenn in einem Extremfall jedoch nur eine Spektralkomponente einen sehr hohen Wert hat, während andere Spektralkomponenten $X(n)$ sehr kleine Werte haben, wird das SFM-Maß einen Wert nahe 0 haben, was ein sehr tonales Signal andeutet.

[0064] Das SFM-Konzept ist sowie andere Merkmalsextraktionskonzepte, um Fingerabdrücke zu erzeugen, sind z. B. in der WO 03/007185 dargestellt.

[0065] In der Identifikations-Phase, die in [Fig. 10](#) dargestellt ist, findet ebenfalls typischerweise dieselbe Merkmalsextraktion **900** wie bei der Trainings-Phase statt. Insbesondere wird der Fingerprint, welcher von dem Audio-Objekt am Audioeingang für einen Zeitabschnitt Δt extrahiert wurde, mit den Referenz-Fingerprints der Fingerprint-Datenbank **902** mittels eines Vergleichers **904** verglichen, wobei der Vergleich typischerweise in der Einrichtung **12** zum Bereitstellen von Identifikationsergebnissen, wie sie anhand von [Fig. 1](#) dargestellt worden ist, enthalten sein wird. Hierauf erhält man bei Feststellung einer Übereinstimmung anhand eines bestimmten Kriteriums ein Erkennungsergebnis für den Zeitabschnitt Δt . Wird also eine Übereinstimmung anhand eines bestimmten Kriteriums festgestellt, so kann der unbekannte Fingerprint, und damit der Ausschnitt aus dem unbekanntem Audio-Objekt, einem Referenzmaterial in der Datenbank, also einer Liste von Identifikationsergebnissen ID_i, ID_{i+1}, \dots mit verschiedenen Zuverlässigkeitswerten zugeordnet werden.

[0066] Erfindungsgemäß wird nunmehr nicht nur ein unbekanntes Audioobjekt am Eingang genau einem Referenz-Audioobjekt in der Referenz-Datenbank zugeordnet, und zwar nur für einen Zeitpunkt Δt , sondern es wird kontinuierlich gearbeitet, ohne Unterbrechung des Datenstroms am Eingang. Erfindungsgemäß wird eine Zuordnung verschiedener Ausschnitte aus Audio-Objekten jeweils zu den korrekten Audio-Objekten aus der Referenzdatenbank vorgenommen. Damit wird eine lückenlose Abfolge, d. h. ein Protokoll, der identifizierten Audio-Objekte am Eingang erhalten.

[0067] Nachfolgend wird anhand der [Fig. 4a](#) bis [Fig. 5d](#) eine besondere Schwierigkeit der kontinuierlichen Analyse eines kontinuierlichen Audiodatenstroms dargestellt. Es muß nämlich das Audio-Objekt in Abschnitte der Länge Δt_x , also in einzelne Blöcke zerlegt werden, um für den Teilabschnitt des Audiodatenstroms eine Zuordnung zu einem Referenzelement in der Datenbank vornehmen zu können. Es kann nun vorkommen, daß diese Zuordnung eines einzelnen Abschnitts des Audiodatenstroms nicht immer eindeutig ist und erst im Zusammenhang mit vorherigen und folgenden Zuordnungen eindeutig wird. Wenn man Einzelzuordnungen vornimmt und diese in einem weiteren Schritt lediglich zusammenfaßt, erhält man fehlerhafte Erkennungsprotokolle, was nachfolgend dargestellt wird.

[0068] [Fig. 4a](#) stellt eine lange Version eines Musikstücks XY dar, das auch durch einen in [Fig. 4a](#) dargestellten langen Fingerprint dargestellt wird, wobei diesem Fingerprint das Identifikationsergebnis ID108

zugeordnet ist. [Fig. 4b](#) zeigt dasselbe für eine Kurzversion des gleichen Musikstücks XY. ID109 verweist somit auf eine Kurzversion des Musikstücks XY, während ID108 auf eine Langversion dieses Musikstücks verweist. Nachdem die Kurzversion kürzer als die Langversion ist, ist auch der Fingerprint in [Fig. 4b](#) kürzer als der Fingerprint in [Fig. 4a](#). In der Weise, in der beide Blöcke untereinander dargestellt sind, enthalten die Musikstücke und damit auch die Fingerprints ID108 und ID109 identisches Audiomaterial bzw. identische Fingerprintdaten. ID109 ist also eine Teilmenge von ID108. Aus [Fig. 4c](#) ist damit ersichtlich, daß die lange Version einen Anfangsabschnitt hat im Zeitraum Δt_0 , der in der kurzen Version nicht vorhanden ist. Im Mittelabschnitt zwischen t_1 bis t_5 sind die Langversion und die Kurzversion identisch, während die lange Version wieder zwischen dem Zeitpunkt t_5 und t_7 einen Musikabschnitt hat, der in der Kurzversion, die durch ID109 identifiziert ist, nicht vorhanden ist.

[0069] Nachfolgend wird anhand der [Fig. 5a](#) bis [Fig. 5d](#) dargestellt, wie bei einfachem Zusammenfassen, also ohne Hypothesenbildung mit den Einzelidentifikationen fehlerhafte Erkennungsprotokolle entstehen können. Es sei angenommen, daß am Eingang des Systems zum Zeitpunkt t_0 das Musikstück ID108 aufgenommen wird. Ferner sei die Datenbank dahingehend wirksam, daß sie die in [Fig. 5a](#) gezeigten Elemente für die Zeitabschnitte Δt_x identifiziert. Es sei darauf hingewiesen, daß die Identifikation in [Fig. 5a](#) prinzipiell richtig ist, obgleich jedoch in den Zeitabschnitten Δt_1 bis Δt_4 sowohl ID108 als auch ID109 ausgegeben werden könnte. Letztendlich ist die Bestimmung der Identifikationsergebnisse in diesen Bereichen zweideutig, da die Datenbank bei nicht vorhandener Störung sowohl ID109 als auch ID108 ausgeben wird und sich aufgrund rechnerischer Unterschiede z. B. immer für den wahrscheinlichsten Wert entscheiden wird, so daß immer aufgrund irgendwelchen Rauschens einer der beiden Identifikationsergebnisse ID108 oder ID109 ein leicht höheres Zuverlässigkeitsmaß haben wird. Im Erkennungsprotokoll, das in [Fig. 5b](#) dargestellt ist, wird somit eine falsche Identifikation dahingehend vorgenommen, daß zu keinem Zeitpunkt das durch ID109 identifizierte Stück gespielt worden ist, sondern daß lediglich das mit ID108 identifizierte Stück abgespielt worden ist.

[0070] Nachfolgend wird anhand der [Fig. 5c](#) und [Fig. 5d](#) eine weitere Alternative dargestellt. Es wird davon ausgegangen, daß die Datenbank die in [Fig. 5c](#) gezeigte Situation ausgibt. Im Erkennungsprotokoll wird wieder eine falsche Zusammenfassung gegeben, nämlich dahingehend, daß ID109 zwischen T_1 und T_5 vorhanden war, während dies natürlich nicht zutreffend ist. Statt dessen wurde die Langversion des Musikstücks, also ID108 von t_0 bis t_7 abgespielt.

[0071] Ferner sind weitere Falscherkennungsprotokolle denkbar, welche durch die Mehrdeutigkeit der Einzelerkennungen für einen Abschnitt des Audiodatenstroms im Zeitraum Δt_x entstehen werden.

[0072] Erfindungsgemäß wird nunmehr auf das in [Fig. 6](#) dargestellte allgemeine Konzept zugegriffen, bei dem die für einen Zeitabschnitt Δt_x erhaltenen Erkennungsergebnisse, also die Ausgangssignale der Einrichtung **12** von [Fig. 1](#), die je nach Implementierung die Einrichtungen **900**, **904**, **902** vereinigen kann, einem Postprocessing unterzogen werden, das im wesentlichen in der Einrichtung zum Bilden von wenigstens zwei Hypothesen und der Einrichtung zum Untersuchen der Hypothesen von [Fig. 1](#) entspricht. Hierauf wird dann unter Verwendung des Postprocessings, also unter Verwendung der bei dem Postprocessing erhaltenen Untersuchungsergebnisse eine Aussage über das Informationssignal in Form einer Erkennungsabfolge bzw. eines Erkennungsprotokolls gezogen.

[0073] In der Postprocessing-Stufe wird die Wahrscheinlichkeit für den Übergang von einem identifizierten Referenzaudioobjekt für den Zeitabschnitt Δt_x zu einem beliebigen anderen Referenzaudioobjekt für den Zeitabschnitt Δt_{x+1} als gleich angenommen. Hieraus werden verschiedene zunächst parallel betrachtete Hypothesen für zusammenhängende Audioabschnitte aus den Einzelerkennungen gebildet. Es ist zu beachten, daß Einzelerkennungen dann zu einer Hypothese zusammengefügt werden, wenn sie sich auf ein und dasselbe Referenzaudiosignal beziehen und zeitkontinuierlich zusammenhängen. Das Erkennungsprotokoll ergibt sich aus einem Zusammenfügen der jeweils wahrscheinlichsten Hypothesen unter Beachtung des zeitlichen Fortschritts. Nachfolgend wird ein bevorzugter Algorithmus detailliert dargestellt.

[0074] Zunächst werden verschiedene Hypothesen für zusammenhängende Audioabschnitte aus den Einzelerkennungen für die Zeitabschnitte Δt_x (mit $x = N, N+1, N+2, \dots$; wobei t_n der Startzeitpunkt für die jeweilige Hypothese darstellt) für je ein erkanntes Referenzaudioobjekt gebildet.

[0075] Einzelerkennungen werden dann zu einer Hypothese zusammengesetzt, wenn die Einzelerkennungen zeitlich kontinuierlich aufeinander folgen.

[0076] Die zeitliche Kontinuität ist ein weiteres Element, das dazu dient, festzustellen, ob eine bereits bestehende Hypothese fortgesetzt wird, oder ob eine neue Hypothese begonnen wird. So sei das Szenario betrachtet, indem ein bestimmtes z. B. Gitarrensolo in einem Stück in der Kurzversion des Stücks ziemlich am Anfang des Stücks kommt und in einer Langversion des Stücks eher in der Mitte des Stücks.

[0077] Bei einem bevorzugten Ausführungsbeispiel gibt die Datenbank, also die Einrichtung zum Bereitstellen von Identifikationsergebnissen, nicht nur eine Fingerabdruckidentifikation aus, sondern auch einen Zeitwert, der sich daraus ergibt, dass der Identifikationsfingerabdruck in der Datenbank eine Länge hat, und der eingegebene (kurze) Fingerabdruck lediglich mit einem Teil des (langen) Fingerabdrucks in der Datenbank übereinstimmt.

[0078] Bei dem vorstehend beschriebenen Szenario würde die Datenbank für das Gitarrensolo vielleicht zwei ID-Ergebnisse liefern (kurze Version und lange Version), jedoch mit zwei unterschiedlichen Zeitindizes. Der Zeitindex für das ID-Ergebnis für die kurze Version ist dabei kleiner als der Zeitindex für die lange Version. Auf der Basis des Zeitindex ist die Einrichtung zum Bilden der Hypothesen nunmehr in der Lage, Hypothesen fortzuführen (wenn eine zeitliche Kontinuität zwischen dem Zeitindex und dem letzten Zeitindex in der Hypothese ist), oder neue Hypothesen zu beginnen, wenn keine Kontinuität in dem aktuell erhaltenen Zeitindex und einem letzten Zeitindex einer Hypothese vorhanden ist.

[0079] Jede zeitliche Diskontinuität bezogen auf ein Referenzaudioobjekt erzeugt eine neue Hypothese, wenn das folgende Element zeitlich einen größeren Abstand als ein festzulegender Zeitabstand T_a ist, oder wenn das folgende Element zeitlich vor dem vorherigen liegt.

[0080] Zur Hypothesenuntersuchung wird für jede Hypothese eine Addition der Konfidenzmaße, also der Zuverlässigkeitswerte bzw. der Maße für die Plausibilität, der Einzelerkennungen vorgenommen.

[0081] Beginnend mit dem Zeitabschnitt Δt_0 wird dann die Hypothese mit dem höchsten Konfidenzmaß als wahr gewertet und in das Erkennungsprotokoll übernommen. Für den nächsten Zeitabschnitt, welcher der ersten Hypothese folgt, wird erneut die Hypothese mit dem höchsten Konfidenzmaß als wahr gewertet und in das Erkennungsprotokoll übernommen usw.

[0082] Für das oben bezeichnete Beispiel ergibt sich somit ein Ablauf, der anhand der [Fig. 7a](#) bis [Fig. 7c](#) dargestellt ist. Für den Zeitabschnitt Δt_0 liefert die Datenbank, wie sie beispielsweise in [Fig. 2](#) dargestellt ist, nur ein Identifikationsergebnis, nämlich ID108, welches eine Wahrscheinlichkeit bzw. ein Zuverlässigkeitsmaß hat, das oberhalb einer Schwelle ist. Im Zeitintervall Δt_1 , also für den Block von Informationseinheiten, die sich über das Zeitintervall Δt_1 erstrecken, liefert die Datenbank zwei Ergebnisse, die ein Zuverlässigkeitsmaß haben, das über einer Schwelle ist. Die zwei Ergebnisse werden auch für die Blöcke zwischen den Zeitpunkten t_2 bis t_5 erhalten. Dann, für den Zeitraum t_5 bis t_7 liefert die Da-

tenbank wiederum lediglich ein einziges Identifikationsergebnis, dessen Zuverlässigkeitsmaß oberhalb einer Schwelle ist.

[0083] Die Einrichtung **14** ([Fig. 1](#)) zum Bilden von wenigstens zwei Hypothesen ist ausgebildet, um zum Zeitpunkt t_0 aufgrund des Identifikationsergebnisses ID108 eine erste Hypothese zu starten, und um zum Zeitpunkt t_1 aufgrund des neu hinzugekommenen Identifikationsergebnisses ID109 eine neue Hypothese, nämlich die Hypothese H2, zu starten.

[0084] Irgendwann nach dem Zeitpunkt t_7 wird dann die in [Fig. 7a](#) gezeigte Hypothesensituation mit den Hypothesen H1 und H2 betrachtet, um dann aufgrund der Untersuchung der Hypothesen, die so ablaufen kann, wie es in [Fig. 7b](#) dargestellt ist, für jede Hypothese die Funktionen für die Konfidenzmaße der Einzelerkennungen, also x_{H1} und x_{H2} zu berechnen.

[0085] Wenn davon ausgegangen wird, daß zwischen t_1 und t_5 die Identifikationsergebnisse ID108 und ID109 mit gleicher Wahrscheinlichkeit auftreten, so wird bei dem in [Fig. 7a](#) gezeigten Ausführungsbeispiel lediglich die erste Hypothese H1 das Rennen machen, da die Hypothese zwar zwischen t_1 und t_5 genauso wahrscheinlich war wie die Hypothese H2, da die Hypothese H1 jedoch im Zeitraum Δt_0 und im Zeitraum Δt_5 und im Zeitraum Δt_6 gilt, also ein Zuverlässigkeitsmaß für eine Einzelerkennung beiträgt, die für die Hypothese H2 nicht gegeben wird. Dies bedeutet somit für das Erkennungsprotokoll den in [Fig. 7c](#) gezeigten richtigen Fall, daß nämlich das mit ID108 bezeichnete Stück vom Zeitpunkt t_0 bis zum Zeitpunkt t_7 abgespielt worden ist.

[0086] Beginnend bei t_0 wird somit die Hypothese H1 ausgewählt, da es bis t_7 keine Hypothese mit einem höheren Konfidenzmaß gibt. Die Hypothese H2 wird verworfen, wobei prinzipiell sämtliche Hypothesen verworfen werden können, die parallel zu einer anderen Hypothese existieren, die als die wahrscheinlichste ausgewählt worden ist.

[0087] Erfindungsgemäß wird somit genau die Sequenz, hier im Beispiel ein Element, nämlich ID108, protokolliert, welche auch tatsächlich am Audioeingang abgespielt wurde.

[0088] Es sei darauf hingewiesen, daß es für die Bestimmung des Endes einer Hypothese verschiedene Möglichkeiten gibt. So kann – unabhängig von der Hypothesensituation – ein Informationsentitätssende z. B. ermittelt werden aus dem Audiosignal selbst, wenn beispielsweise eine Pause mit einer bestimmten Minimallänge auftritt. Nachdem dieses Kriterium jedoch dann nicht wirkt, wenn zwischen zwei Informationsentitäten überblendet („Fading“) wird, oder wenn zwei Stücke derart schnell aufeinanderfolgen,

daß keine merkbare Pause zu finden ist, wird es bevorzugt, aufgrund der in der Vergangenheit betrachteten Hypothesen ein Informationsentitätsende zu bestimmen. Dies kann beispielsweise dahingehend stattfinden, daß eine Hypothese dann als beendet angesehen wird, wenn z. B. zwei oder mehr Blöcke kein Identifikationsergebnis mehr mit einem Zuverlässigkeitswert oberhalb einer bestimmten Minimalschwelle der Einrichtung **14** zum Bilden von Hypothesen zugeführt werden. Alternativ kann auch, beispielsweise für den in **Fig. 3** gezeigten Fall einfach damit begonnen werden, zu irgendeinem Zeitpunkt in die Vergangenheit gerichtet die Werte der Hypothesen für eine vorbestimmte Anzahl von Blöcken aufzuaddieren, um dann zu sehen, welche Hypothese am Ende, also nach einer bestimmten Anzahl von z. B. 20 Blöcken für bestimmte Blöcke den höchsten Wert hatte und damit überlebt hat und die anderen Hypothesen „ausgestochen“ hat. Dies würde bei dem in **Fig. 3** gezeigten Beispiel bedeuten, daß die Hypothesen, daß die Informationsentität ID1 oder ID2 oder ID3 ist, auch für die Zeitabschnitte $\Delta t7$ und $\Delta t8$ weitergeführt werden würden, wobei dies jedoch nichts an der Erkennung von ID1 ändern würde, da neue Hypothesen, also die Hypothese für ID108, ID109, ID4 und ID8 erst wesentlich später, also für Blöcke von $\Delta t7$ und $\Delta t8$ oder darüber begonnen werden und damit erst viel später oder überhaupt nicht derart hohe kombinierte Zuverlässigkeitswerte erreichen.

[0089] Aus dem vorstehenden wird ersichtlich, daß das Ende einer Hypothese nicht unbedingt aktiv bestimmt werden muß, sondern daß sich dieses Ende aus der Analyse der Vergangenheit, also der begonnenen Hypothesen automatisch ergeben kann. Vorzugsweise wird daher immer dann, wenn ein neues Identifikationsergebnis mit einem Zuverlässigkeitsmaß oberhalb einer Signifikanzschwelle in Erscheinung tritt, eine neue Hypothese begonnen, wobei dann irgendwann in die Vergangenheit geblickt wird, um zu sehen, welche Hypothese für einen bestimmten Zeitraum überlebt, wobei hierfür nicht explizit ein Ende einer Hypothese bestimmt werden muß, da es sich automatisch selbst ergibt.

[0090] Abhängig von der Gegebenheit, kann das erfindungsgemäße Verfahren 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 Verfahren ausgeführt wird. Allgemein besteht die Erfindung somit auch in einem Computer-Programm-Produkt mit einem auf einem maschinenlesbaren Träger gespeicherten Programmcode zur Durchführung des erfindungsgemäßen Verfahrens, wenn das Computer-Programm-Produkt auf einem Rechner abläuft. In anderen Worten ausge-

drückt, kann die Erfindung somit als ein Computer-Programm mit einem Programmcode zur Durchführung des Verfahrens realisiert werden, wenn das Computer-Programm auf einem Computer abläuft.

Patentansprüche

1. Vorrichtung zum Analysieren eines Informationssignals, das eine Folge (**802**) von Blöcken von Informationseinheiten (**804**) aufweist, wobei eine Mehrzahl von aufeinanderfolgenden Blöcken der Folge von Blöcken eine Informationsentität (**806**) darstellt, unter Verwendung einer Folge von Fingerabdrücken (FAi) für die Folge von Blöcken, so daß die Folge von Blöcken durch die Folge von Fingerabdrücken dargestellt ist, mit folgenden Merkmalen:

einer Einrichtung (**12**) zum Bereitstellen von Identifikationsergebnissen (IDi) für aufeinanderfolgende Fingerabdrücke, wobei ein Identifikationsergebnis eine Zugehörigkeit eines Blocks von Informationseinheiten zu einer vorbestimmten Informationsentität darstellt, und wobei es für jedes Identifikationsergebnis ein Zuverlässigkeitsmaß gibt, wobei die Einrichtung (**12**) zum Bereitstellen ausgebildet ist, um für einen ersten Fingerabdruck ein erstes Identifikationsergebnis zu erzeugen, und um für einen folgenden Block ein zweites Identifikationsergebnis zu erzeugen, das sich von dem ersten Identifikationsergebnis unterscheidet;

einer Einrichtung (**14**) zum Bilden von wenigstens zwei Hypothesen aus den Identifikationsergebnissen für die aufeinanderfolgenden Fingerabdrücke, wobei eine erste Hypothese eine Annahme für die Zugehörigkeit der Folge von Blöcken zu einer ersten Informationsentität ist, und wobei eine zweite Hypothese eine Annahme für die Zugehörigkeit der Folge von Blöcken zu einer zweiten Informationsentität ist, wobei die Einrichtung (**14**) zum Bilden ausgebildet ist, um ansprechend auf das erste Identifikationsergebnis die erste Hypothese zu beginnen oder die bereits bestehende erste Hypothese weiterzuführen und um ansprechend auf das zweite Identifikationsergebnis die zweite Hypothese zu beginnen oder die bereits bestehende zweite Hypothese weiterzuführen;

einer Einrichtung (**16**) zum Untersuchen der wenigstens zwei Hypothesen durch Zusammenfassen der Zuverlässigkeitsmaße der Hypothesen, um ein Untersuchungsergebnis (**18**) zu erhalten; und einer Einrichtung (**20**) zum Treffen einer Aussage über das Informationssignal basierend auf dem Untersuchungsergebnis.

2. Vorrichtung nach Anspruch 1, bei der die Einrichtung (**16**) zum Untersuchen ausgebildet ist, um die Hypothesen im Hinblick auf für die Hypothesen gültige Wahrscheinlichkeitsinformationen zu untersuchen.

3. Vorrichtung nach Anspruch 1 oder 2, bei der die Einrichtung (**20**) zum Treffen einer Aussage aus-

gebildet ist, um zu bestimmen, daß die Folge von Blöcken eine Informationsentität darstellt, die eine Hypothese hat, die am wahrscheinlichsten ist, oder daß eine Informationsentität mit dem Fingerabdruck beendet ist, der als zeitlich letzter zu der wahrscheinlichsten Hypothese beiträgt, oder daß eine Informationsentität in dem Informationssignal vorhanden ist oder nicht.

4. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (12) zum Bereitstellen ausgebildet ist, um für einen Fingerabdruck zwei unterschiedliche Identifikationsergebnisse zu erzeugen.

5. Vorrichtung nach Anspruch 4, bei der die Einrichtung (12) zum Bereitstellen ausgebildet ist, um für jeden der zwei unterschiedlichen Identifikationsergebnisse ein Zuverlässigkeitsmaß zu erzeugen.

6. Vorrichtung nach Anspruch 4 oder 5, bei der die Einrichtung (14) zum Bilden ausgebildet ist, um der ersten Hypothese ein erstes der zwei Identifikationsergebnisse und der zweiten Hypothese ein zweites der zwei Identifikationsergebnisse zuzuordnen.

7. Vorrichtung nach einem der Ansprüche 3 bis 6, bei der die Einrichtung (16) zum Untersuchen ausgebildet ist, um die Hypothese zu bestimmen, die ein höheres zusammengefaßtes Zuverlässigkeitsmaß hat.

8. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (14) zum Bilden ausgebildet ist, um die erste oder zweite Hypothese zu beenden, wenn eine vorbestimmte Anzahl von Blöcken weder ein Identifikationsergebnis, das auf die erste Informationsentität hinweist, noch ein Identifikationsergebnis, das auf die zweite Informationsentität hinweist, erhalten wird.

9. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (14) zum Bilden ausgebildet ist, um die erste oder zweite Hypothese zu beenden, wenn ein detektiertes Ereignis in dem Informationssignal auftritt.

10. Vorrichtung nach Anspruch 9, bei der ein Ereignisdetektor vorhanden ist, der ausgebildet ist, um einen Energiepegel in einem Block von Informationseinheiten, der unter einem Schwellenpegel liegt, als das Ereignis zu detektieren.

11. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (12) zum Bereitstellen ausgebildet ist, um für jeden Fingerabdruck nur das zuverlässigste Identifikationsergebnis ohne oder mit Zuverlässigkeitsmaß auszugeben, um für einen Fingerabdruck eine vorbestimmte Anzahl von zuverlässigsten Fingerabdrücken jeweils mit oder

ohne Zuverlässigkeitsmaß auszugeben, oder um für einen Fingerabdruck lediglich die Identifikationsergebnisse mit oder ohne Zuverlässigkeitsmaße auszugeben, die ein Zuverlässigkeitsmaß oberhalb einer Schwelle haben.

12. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (16) zum Untersuchen ausgebildet ist, um explizite oder implizite Zuverlässigkeitsmaße, die zu einer Hypothese gehören, zusammenzuaddieren, um ein zusammengefaßtes Zuverlässigkeitsmaß zu erhalten.

13. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (12) zum Bereitstellen ausgebildet ist, um mit einem Fingerabdruck eine Suche in einer Datenbank, in der Fingerabdrücke von Referenz-Informationsentitäten gespeichert sind, durchzuführen, und um eine Anzahl von Identifikationsergebnissen sowie ein Abstandsmaß für jedes Identifikationsergebnis als Hinweis auf ein Zuverlässigkeitsmaß für jedes Identifikationsergebnis zu liefern.

14. Vorrichtung nach Anspruch 13, bei der die Einrichtung (12) zum Bereitstellen ausgebildet ist, um für jedes Identifikationsergebnis, für das nicht bereits eine Hypothese existiert, eine neue Hypothese zu beginnen, wenn ein Abstandsmaß für das Identifikationsergebnis eine Beziehung zu einer Schwelle hat, die auf einen kleineren Abstand als einen Schwellenabstand hinweist.

15. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (16) zum Untersuchen ausgebildet ist, um ansprechend auf eine Bestimmung alle Hypothesen für die aufeinanderfolgenden Fingerabdrücke zu beenden, die für die Fingerabdrücke gebildet worden sind, die von der wahrscheinlichsten Hypothese erfaßt sind.

16. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der das Informationssignal ein Audiosignal umfaßt, bei der die Informationseinheit Audioabstastwerte im Zeit- oder Frequenzbereich sind, und bei der eine Informationsentität ein Musikstück, eine gesprochene Sequenz oder einen Geräuschabschnitt umfaßt.

17. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der ein Fingerabdruck für einen Block durch eine Zeit-Frequenz-Umsetzung und/oder durch Berechnung eines spektralen Flachheitsmaßes für ein Ergebnis der Zeit-Frequenz-Umsetzung ermittelt wird.

18. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der ein Fingerabdruck für einen Block so erzeugt wird, daß der Fingerabdruck eine Daten-

menge hat, die kleiner ist als eine Datenmenge des Blocks.

19. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung (**12**) zum Bereitstellen von Identifikationsergebnissen ausgebildet ist, um neben einem Identifikationsergebnis auch einen neuen Zeitindex für das Identifikationsergebnis zu liefern, und bei der die Einrichtung (**14**) zum Bilden von Hypothesen ausgebildet ist, um eine Hypothese weiterzuführen, wenn eine Kontinuität zwischen einem aktuellsten Zeitindex in der Hypothese und dem neuen Zeitindex existiert, oder um eine Hypothese zu starten, wenn die Kontinuität nicht existiert.

20. Verfahren zum Analysieren eines Informationssignals, das eine Folge (**802**) von Blöcken von Informationseinheiten (**804**) aufweist, wobei eine Mehrzahl von aufeinanderfolgenden Blöcken der Folge von Blöcken eine Informationsentität (**806**) darstellt, unter Verwendung einer Folge von Fingerabdrücken (FAi) für die Folge von Blöcken, so daß die Folge von Blöcken durch die Folge von Fingerabdrücken dargestellt ist, mit folgenden Schritten:

Bereitstellen (**12**) von Identifikationsergebnissen (IDi) für aufeinanderfolgende Fingerabdrücke, wobei ein Identifikationsergebnis eine Zugehörigkeit eines Blocks von Informationseinheiten zu einer vorbestimmten Informationsentität darstellt, und wobei es für jedes Identifikationsergebnis ein Zuverlässigkeitsmaß gibt, wobei im Schritt des Bereitstellens für einen ersten Fingerabdruck ein erstes Identifikationsergebnis und für einen folgenden Block ein zweites Identifikationsergebnis erzeugt werden, das sich von dem ersten Identifikationsergebnis unterscheidet;

Bilden (**14**) von wenigstens zwei Hypothesen aus den Identifikationsergebnissen für die aufeinanderfolgenden Fingerabdrücke, wobei eine erste Hypothese eine Annahme für die Zugehörigkeit der Folge von Blöcken zu einer ersten Informationsentität ist, und wobei die zweite Hypothese eine Annahme für eine Zugehörigkeit der Folge von Blöcken zu einer zweiten Informationsentität ist, wobei der Schritt des Bildens folgende Schritte aufweist:

ansprechend auf das erste Identifikationsergebnis, Beginnen der ersten Hypothese oder Weiterführen der bereits bestehenden ersten Hypothese, und, ansprechend auf das zweite Identifikationsergebnis, Beginnen der zweiten Hypothese oder Weiterführen der bereits bestehenden zweiten Hypothese;

Untersuchen (**16**) der wenigstens zwei Hypothesen durch Zusammenfassen der Zuverlässigkeitsmaße der Hypothesen, um ein Untersuchungsergebnis (**18**) zu erhalten;

und

Treffen (**20**) einer Aussage über das Informationssignal basierend auf dem Untersuchungsergebnis.

21. Computerprogramm mit einem Programm-

code zum Durchführen eines Verfahrens gemäß Patentanspruch 20, wenn das Programm auf einem Computer abläuft.

Es folgen 7 Blatt Zeichnungen

Anhängende Zeichnungen

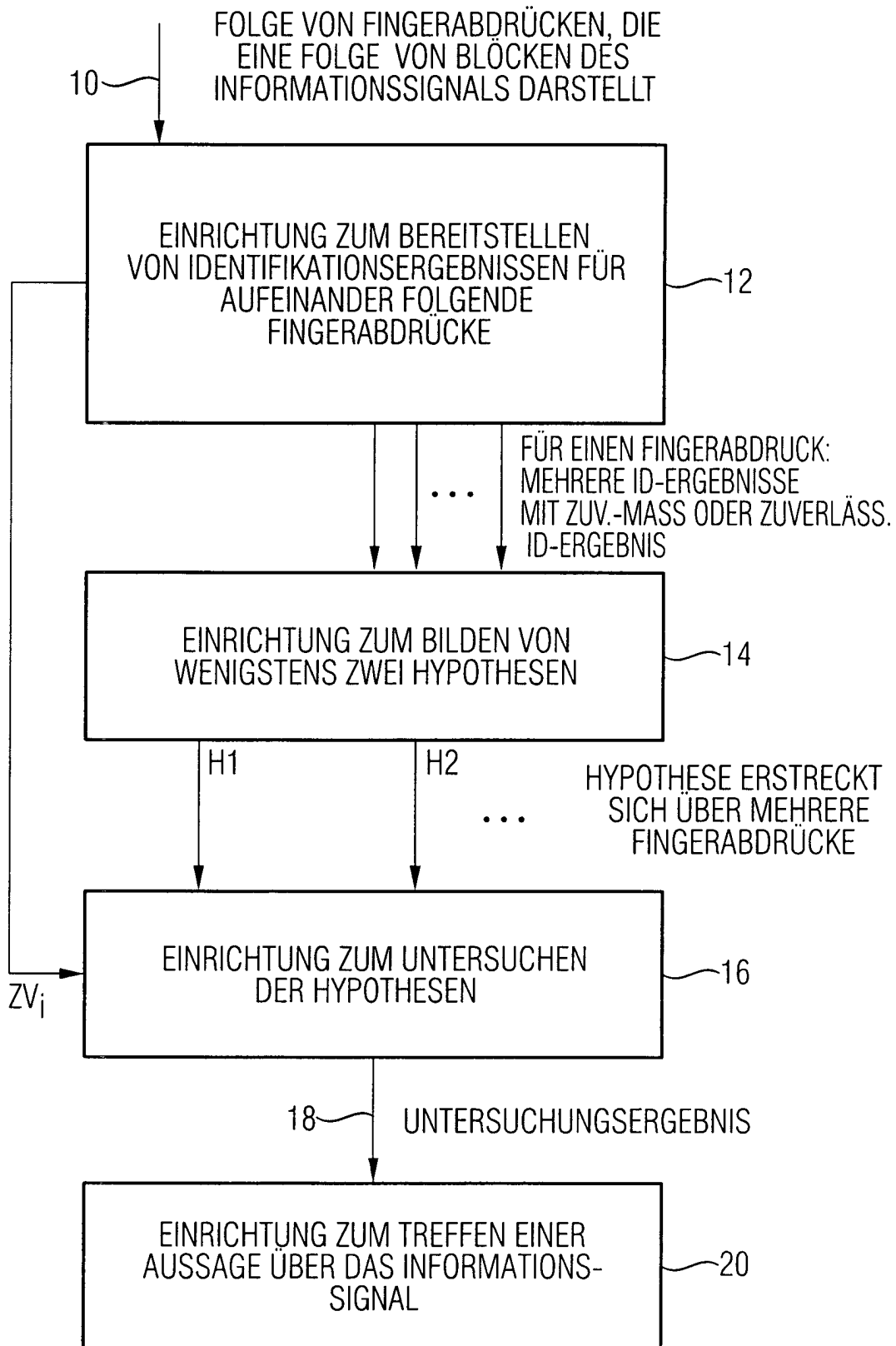


FIG. 1

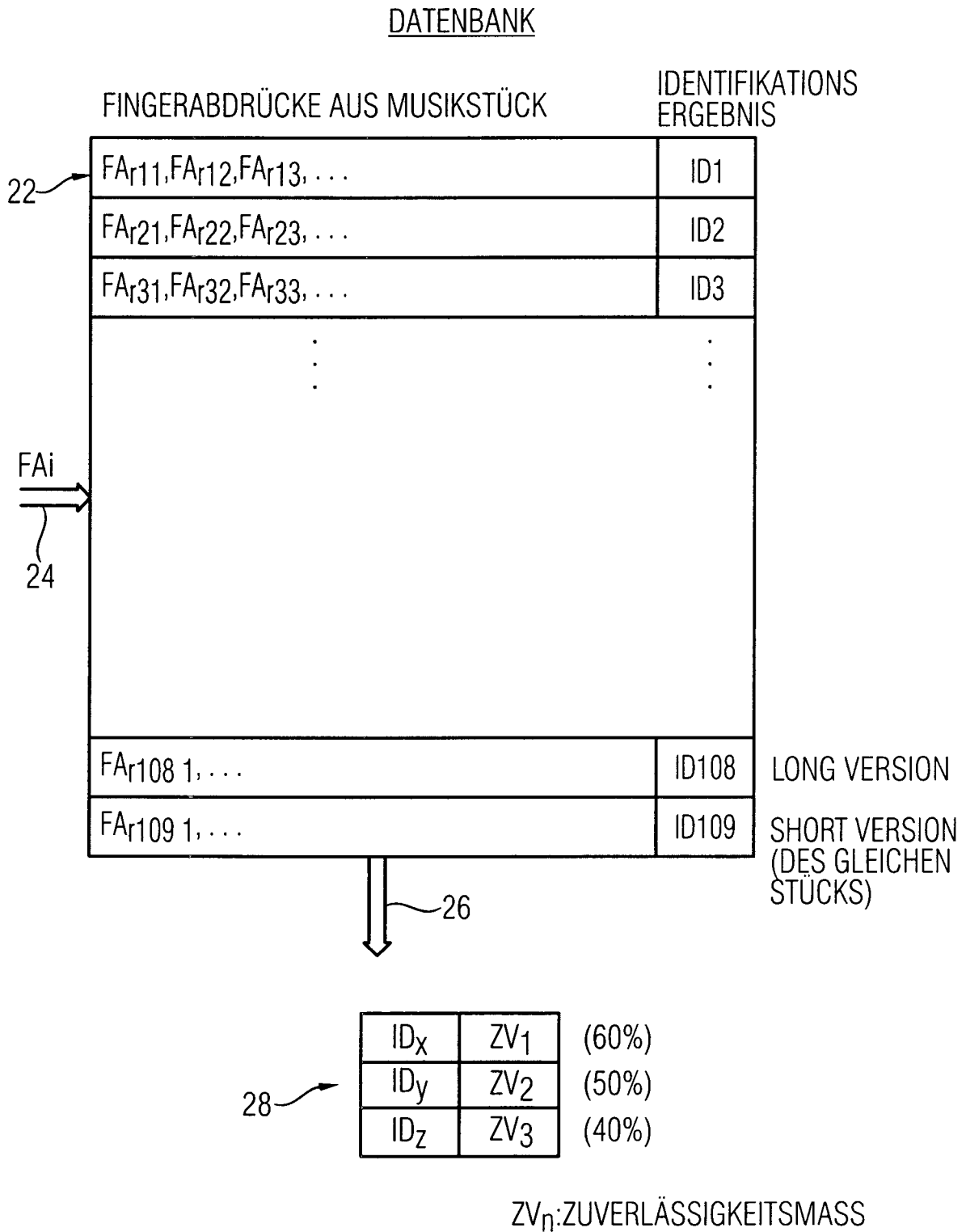


FIG. 2

	Δt_1	Δt_2	Δt_3	Δt_4	Δt_5	Δt_6	Δt_7	Δt_8
	FA1	FA2	FA3	FA4	FA5	FA6	FA7	FA8
	ID1 40%	ID1 60	ID1 80	ID1 60	ID1 90	ID1 90	ID ₁₀₈ 90	ID ₁₀₈ 80
	ID2 60	ID2 70	ID2 30	ID2 40	ID2 20	ID2 10	ID ₁₀₉ 85	ID ₁₀₉ 70
	ID3 30	ID3 20	ID3 30	ID3 40	ID3 10	ID3 5	ID ₄ 10	ID ₈ 10
H1	ID1	ID1	ID1	ID1	ID1	ID1	ID108	ID108
H2	ID2	ID2	ID2	ID2	ID2	ID2	ID109	ID109
H3	ID3	ID3	ID3	ID3	ID3	ID3	ID4	ID4
H4								ID8

FÜR Δt_1 - Δt_6

Σ ID1: 420

Σ ID2: 230

Σ ID3: 135

MUSIKSTÜCK-
ÜBERGANG

⇒ AUSSAGE: -INFORMATIONSENTITÄT VON Δt_1 - Δt_6 IST
 MUSIKSTÜCK ID1; H1 WAR RICHTIG; UND/ODER
 - Δt_6 IST ENDE DER 1. INFORMATIONSENTITÄT
 BZW. Δt_7 IST BEGINN DER 2. INFORMATIONSENTITÄT; UND/ODER
 -ID1 IST IN INFORMATIONSSIGNAL ENTHALTEN

FIG. 3

ID 108 IST EINE LANGE VERSION EINES MUSIKSTÜCKES XY (LÄNGE Z.B. 7:32, SYMBOLISCH DURCH FOLGENDEN FINGERPRINT DARGESTELLT:



FIG. 4A

ID109 IST EINE KURZVERSION DES GLEICHEN MUSIKSTÜCKES XY (Z.B. EINE SOG. "RADIO-VERSION" ODER "RADIO-EDIT", LÄNGE Z.B. 4:38), SYMBOLISCH DURCH FOLGENDEN FINGERPRINT DARGESTELLT:

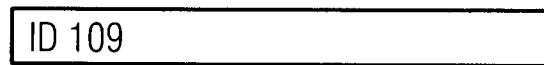


FIG. 4B

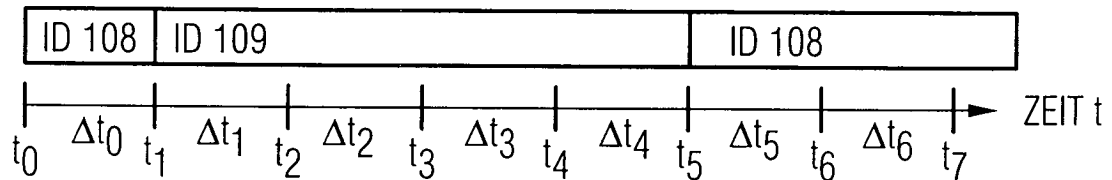


FIG. 4C

IDENTIFIZIERTE ELEMENTE FÜR DIE ZEITABSCHNITTE Δt_x (MIT $x=0,1,2,3,4,5,6$)

ID 108	ID 109	ID 108	ID 109	ID 109	ID 108	ID 108
Δt_0	Δt_1	Δt_2	Δt_3	Δt_4	Δt_5	Δt_6

FIG. 5A

ZUSAMMENFASSUNG IM ERKENNUNGSPROTOKOLL:

1. t_0 BIS t_1 : ID 108 (Δt_0)
2. t_1 BIS t_2 : ID 109 (Δt_1)
3. t_2 BIS t_3 : ID 108 (Δt_2)
4. t_3 BIS t_5 : ID 109 ($\Delta t_3, \Delta t_4$)
5. t_5 BIS t_7 : ID 108 ($\Delta t_5, \Delta t_6$)

FIG. 5B

IDENTIFIZIERTE ELEMENTE FÜR DIE ZEITABSCHNITTE Δt_x (MIT $x=0,1,2,3,4,5,6$)

ID 108	ID 109	ID 109	ID 109	ID 109	ID 108	ID 108
--------	--------	--------	--------	--------	--------	--------

FIG. 5C

ZUSAMMENFASSUNG IM ERKENNUNGSPROTOKOLL:

1. t_0 BIS t_1 : ID 108 (Δt_0)
2. t_1 BIS t_5 : ID 109 ($\Delta t_1 - \Delta t_4$)
3. t_5 BIS t_7 : ID 108 ($\Delta t_5 - \Delta t_6$)

FIG. 5D

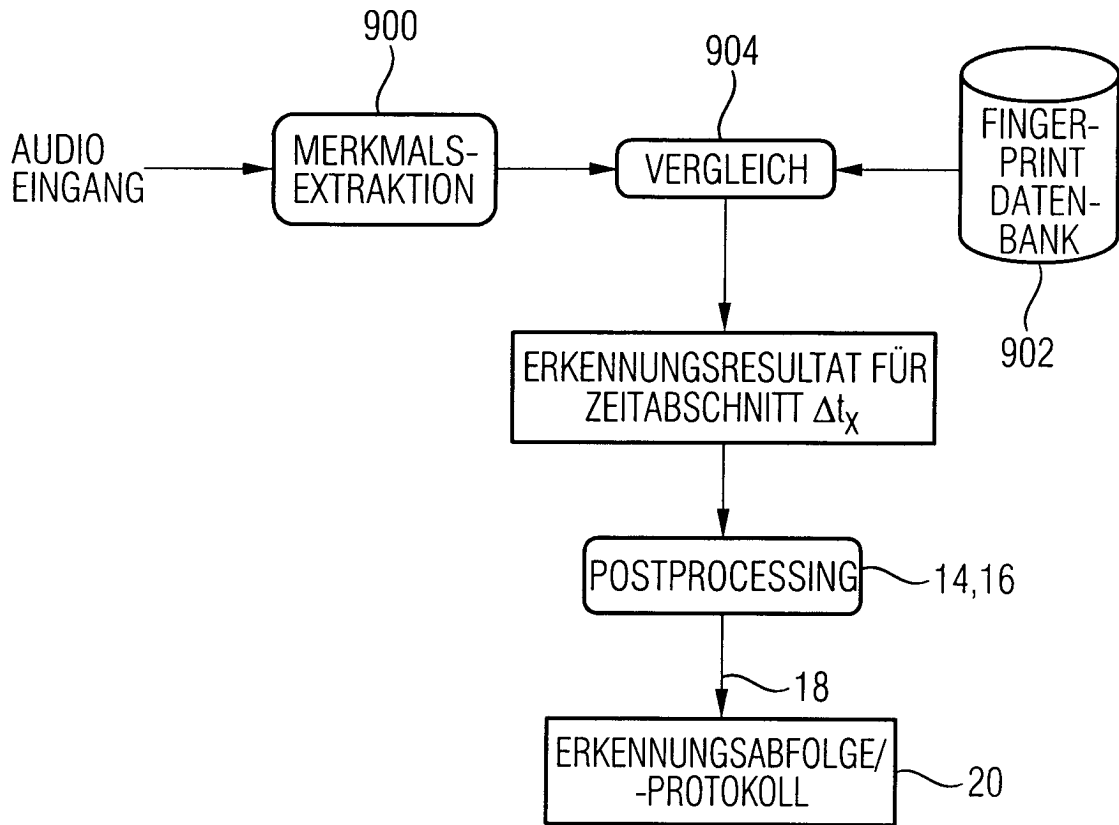


FIG 6

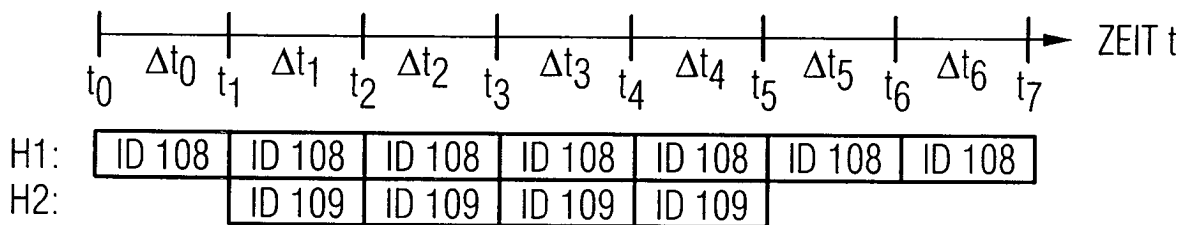


FIG 7A

HYPOTHESE 1 (H1): t_0 BIS t_7 : ID 108, KONFIDENZMASS $x_{H1}\%$
 HYPOTHESE 2 (H2): t_1 BIS t_5 : ID 109, KONFIDENZMASS $x_{H2}\%$
 MIT $x_{H1}\% > x_{H2}\%$, WOBEI x_{H1} UND x_{H2} JE EINE FUNKTION DER KONFIDENZMASSE DER EINZELERKENNUNGEN SIND.

FIG 7B

DIES BEDEUTET FÜR DAS ERKENNUNGSPROTOKOLL:

1. t_0 BIS t_7 : ID 108 (Δt_0 - Δt_6)

FIG 7C

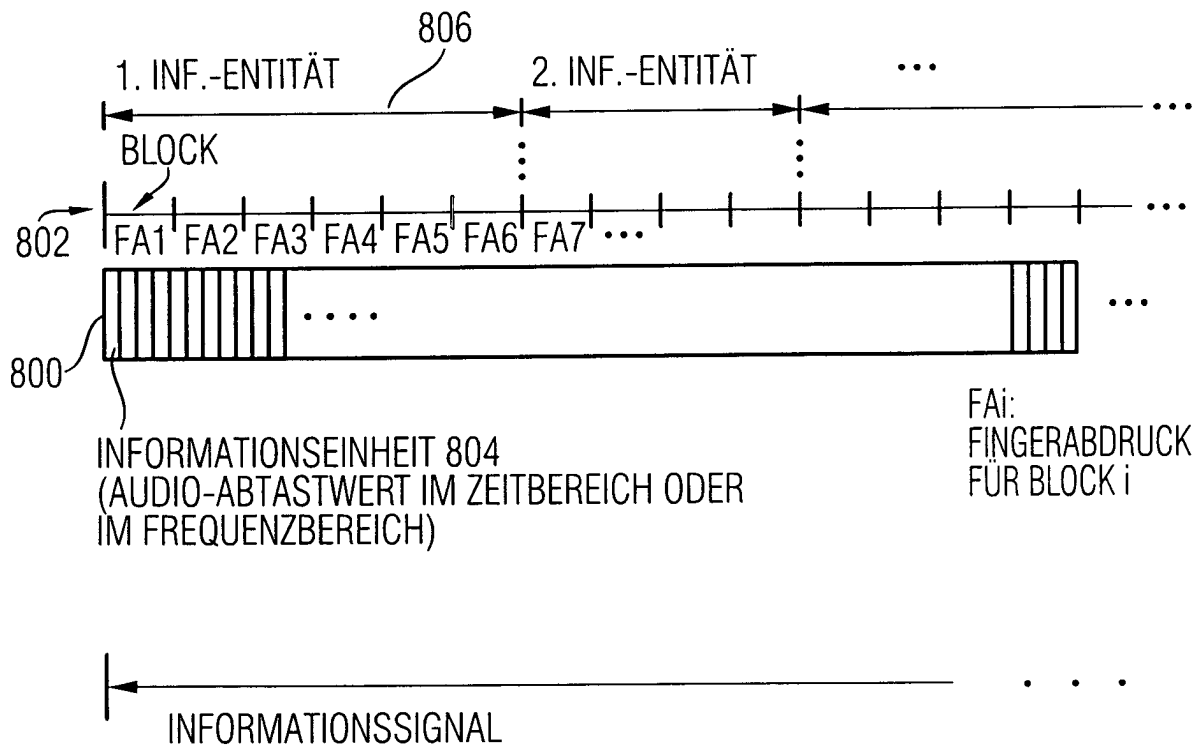


FIG. 8

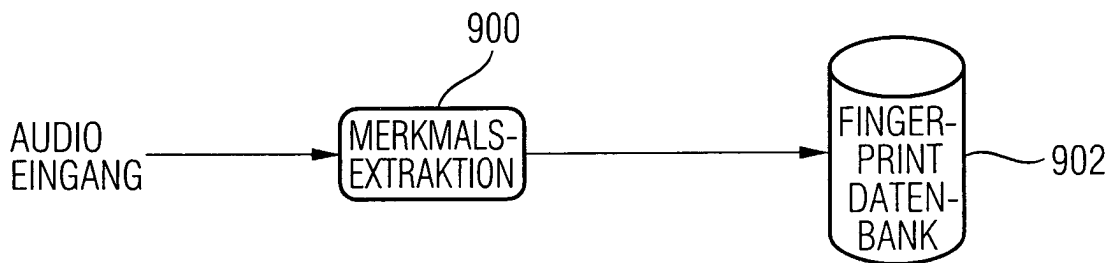


FIG. 9

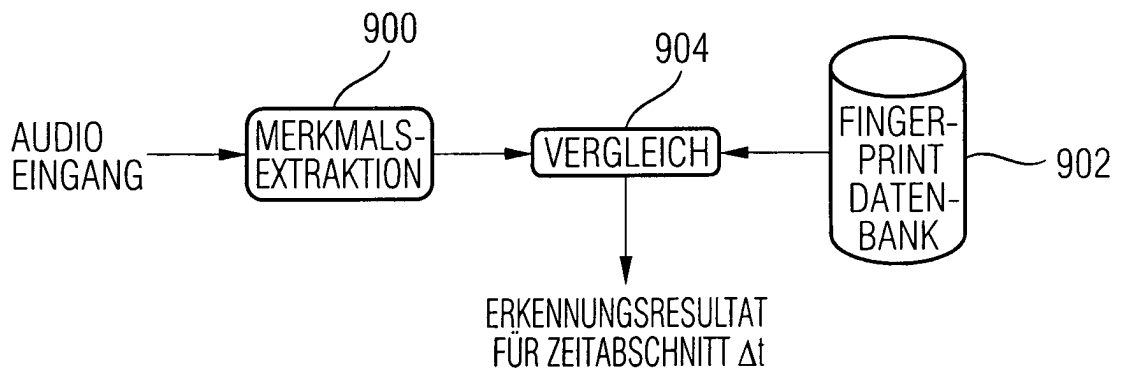


FIG. 10