



(12) **EUROPÄISCHE PATENTANMELDUNG**

(43) Veröffentlichungstag:
04.03.2009 Patentblatt 2009/10

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

(21) Anmeldenummer: **07115473.6**

(22) Anmeldetag: **31.08.2007**

(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 MT NL PL PT RO SE SI SK TR
Benannte Erstreckungsstaaten:
AL BA HR MK RS

(72) Erfinder: **Schlosser, Markus**
30177 Hannover (DE)

(74) Vertreter: **Lindemann, Robert**
Deutsche Thomson OHG
European Patent Operations
Karl-Wiechert-Allee 74
30625 Hannover (DE)

(71) Anmelder: **Deutsche Thomson OHG**
30625 Hannover (DE)

(54) **Verfahren zum Erkennen eines akustischen Ereignisses in einem Audio-Signal**

(57) Ein erfindungsgemäßes Verfahren zum Erkennen eines akustischen Ereignisses eines Audio-Signals (S) weist zwei Stufen (A, B) auf. In der ersten Stufe (A)

werden mögliche Kandidaten (X) ausgewählt und in der zweiten Stufe (B) wird jedem der möglichen Kandidaten (X) ein Konfidenzwert (W) zugeordnet.

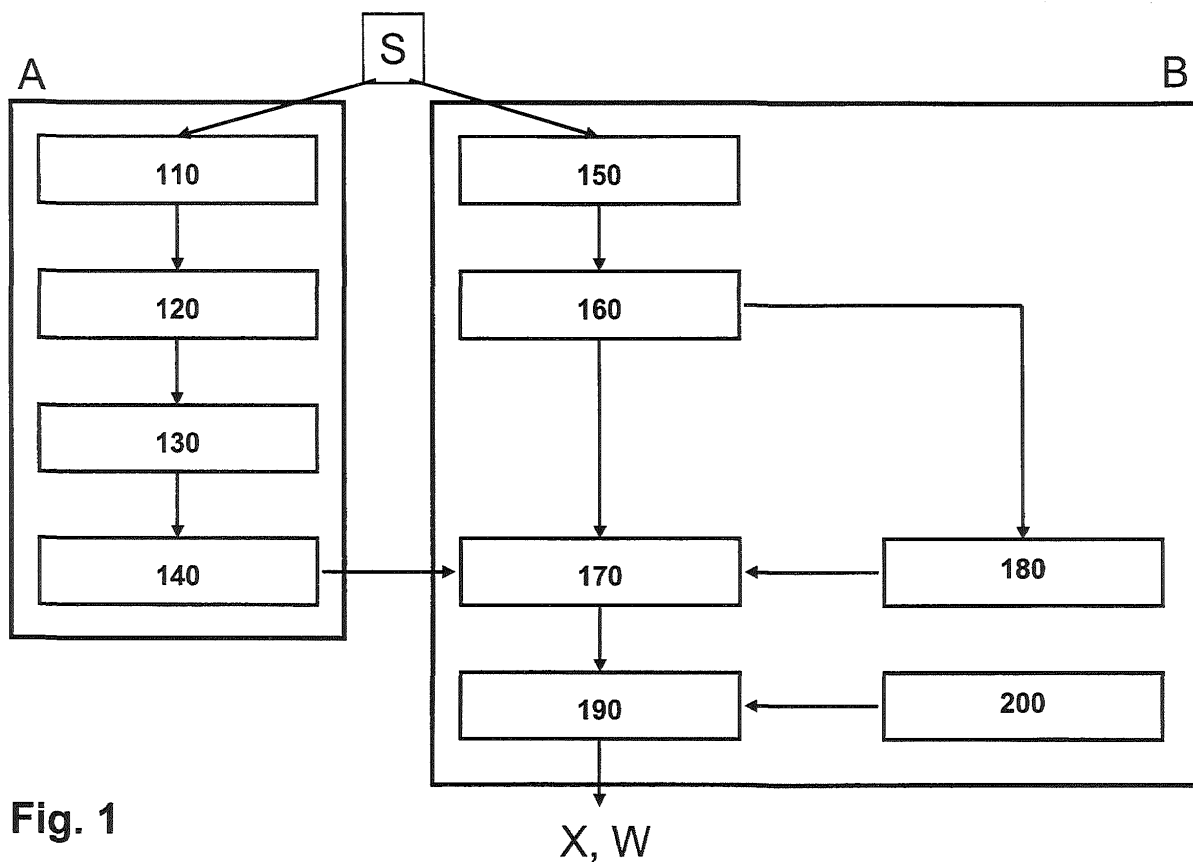


Fig. 1

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zum Erkennen eines akustischen Ereignisses in einem Audio-Signal.

[0002] Es gibt viele Anwendungen, in denen das Erkennen eines akustischen Ereignisses in einem Audio-Signal benötigt wird. Ein Beispiel ist das Erkennen einer Klappe zur Synchronisation von Audio- und Videosignalen. Das Audiosignal einer Klappe gehört als perkusives Signal zu den transienten Signalen. Die Synchronisation von Audio- und Videosignalen wird u. a. bei der Herstellung und Übertragung von Filmen z.B. für Nachrichten benötigt, die möglichst schnell zur Verfügung stehen sollten.

[0003] Ein Verfahren zum Erkennen akustischer Ereignisse in Audio-Signalen ist in der EP 1 465 192 A1 beschrieben. Das Verfahren umfasst eine Stufe, in der eine beliebige Kombination verschiedene Schritte zu einer Klassifikation des Audio-Signals in: erkanntes Ereignis oder kein Ereignis erkannt eingesetzt wird. Die ausgewählten Schritte zur Klassifikation werden an dem gesamten ggf. bearbeiteten Audio-Signal durchgeführt.

[0004] Weitere Verfahren zum Erkennen akustischer Ereignisse sind z.B. aus der US 5 884 260 und aus der US 2005/0199064 A1 bekannt.

[0005] Aufgabe der Erfindung ist es, ein Verfahren zum Erkennen eines akustischen Ereignisses zu entwickeln, das ein schnelles Erkennen des Ereignisses ermöglicht.

[0006] Die Aufgabe ist durch die Merkmale des Anspruchs 1 gelöst. Vorteilhafte Ausführungsformen der Erfindung sind in den Unteransprüchen beschrieben.

[0007] Ein erfindungsgemäßes Verfahren zum Erkennen eines akustischen Ereignisses in einem Audio-Signal, z. B. in einer wav-Datei, weist zwei Stufen auf. In der ersten Stufe werden mögliche Kandidaten ausgewählt und in der zweiten Stufe wird jedem der möglichen Kandidaten ein Konfidenzwert zugeordnet.

[0008] Eine Aufteilung des Erkennungsverfahrens in zwei Stufen, nämlich zunächst eine erste Auswahl möglicher Kandidaten in der ersten Stufe und anschließend eine genauere Überprüfung der möglichen Kandidaten in der zweiten Stufe ermöglicht es, die auszuwertende Datenmenge erheblich gegenüber Verfahren zu verringern, bei denen die Überprüfung der Kandidaten ohne Vorauswahl durchgeführt wird.

[0009] Der Konfidenzwert ist ein Maß für die Wahrscheinlichkeit, dass es sich um das gesuchte Ereignis handelt. Die Zuordnung eines Konfidenzwertes zu jedem möglichen Kandidaten ermöglicht es einem Operator, bei Bestimmung der endgültigen Kandidaten, zunächst die möglichen Kandidaten mit den höchsten Konfidenzwerten zu sichten und die Suche zu beenden, sobald die gesuchten Kandidaten gefunden sind. Mögliche, falsche Kandidaten mit ähnlichen Eigenschaften wie die gesuchten Ereignisse d.h. und mit zwar hohen jedoch etwas geringeren Konfidenzwerten als die gesuchten Ereignisse können außer Betracht gelassen werden.

[0010] Zur Auswahl der möglichen Kandidaten weist die erste Stufe vorzugsweise folgende Schritte auf: Anwendung eines ersten Hoch-Pass-Filters auf das Audio-Signal, Berechnung einer Energieeinhüllenden im Zeitbereich aus dem gefilterten Audio-Signal, Berechnung einer Ableitung aus der Energieeinhüllenden und Bestimmung von möglichen Kandidaten aus Ereignissen, deren Ableitung der Energieeinhüllenden über einem vorbestimmten Schwellenwert liegt. Dies ist ein einfaches Verfahren zur Auswahl der möglichen Kandidaten.

[0011] Bevorzugt weist die zweite Stufe für jeden möglichen Kandidaten folgende Schritte auf: Auswertung einer oder mehrere Größen des möglichen Kandidaten und Zuordnung eines gemeinsamen Konfidenzwertes mit Hilfe einer Bewertung der Größen.

[0012] Dabei werden vorzugsweise eine oder mehrerer der folgenden Größen jedes möglichen Kandidaten ausgewertet:

- Energieanstieg, d.h. der Maximalwert der Ableitung,
- Höhe und Position des gemessenen Maximums,
- Steigung und Abweichung von einer an den Energieabfall der Energieeinhüllenden angepassten Kurve,
- Differenz zwischen einem gemessenen Maximum und einem aus der Kurve vorausgesagten Maximum,
- Dauer des möglichen Kandidaten,
- Dauer einer Stilleperiode vor dem möglichen Kandidaten und Dauer einer Stilleperiode nach dem möglichen Kandidaten, und
- Zeitpunkt des Auftretens des möglichen Kandidaten.

[0013] Bevorzugt weist die zweite Stufe für jeden möglichen Kandidaten für die Auswertung der Größen folgende Schritte auf:

Anwendung eines zweiten Hoch-Pass-Filters auf das Audio-Signal und Berechnung einer Energieeinhüllenden im Zeitbereich aus dem gefilterten Audio-Signal und/oder Bestimmung eines Geräuschbereichs des Audio-Signals.

[0014] In einer Ausführungsform der Erfindung umfasst die Bestimmung des Geräuschbereichs eine Bestimmung eines Grundrauschens und/oder eines Aufnahmepegels. Bevorzugt wird dabei die in der zweiten Stufe berechnete Energieeinhüllende verwendet.

[0015] Bevorzugt weist die zweite Stufe für jeden möglichen Kandidaten für die Bewertung einer oder mehrerer der ausgewerteten Größen jeweils folgende Schritte auf: Bestimmung eines Wahrscheinlichkeitsverhältnisses und/oder eines Gewichtungsfaktor. Bevorzugt werden die Wahrscheinlichkeitsverhältnisse und/oder die Gewichtungsfaktoren der ausgewerteten Größen bei der Zuordnung des gemeinsamen Konfidenzwertes zusammengefasst.

[0016] In einer Ausführungsform der Erfindung erfolgt bei der Zuordnung des gemeinsamen Konfidenzwertes eine Addition der Logarithmen der durch die Gewichtungsfaktoren gewichteten Wahrscheinlichkeitsverhältnisse der ausgewählten Größen.

[0017] Bevorzugt werden die Gewichtungsfaktoren einer oder mehrerer der ausgewerteten Größen jeweils aus Korrelationkoeffizienten für paarweise Korrelationen der ausgewerteten Größen berechnet.

[0018] Bevorzugt werden bei der Bestimmung der Wahrscheinlichkeitsverhältnisse eine oder mehrere Zusatzinformationen über das akustische Ereignis berücksichtigt.

[0019] In einer Ausführungsform der Erfindung weist die zweite Stufe für jeden möglichen Kandidaten alternativ oder zusätzlich folgenden Schritt auf:

Spracherkennung eines auf das akustische Ereignis hinweisenden Textes.

[0020] Das erfindungsgemäße Verfahren wird bevorzugt zum Erkennen von Klappen bei der Synchronisation des Audio-Signals mit einem entsprechenden Video-Signal verwendet.

[0021] Die Erfindung ist anhand eines in der Zeichnung schematisch dargestellten Beispiels weiter erläutert.

[0022] Es zeigen:

Figur 1 eine Block-Diagramm eines erfindungsgemäßen Beispiel und
Figur 2 eine Darstellung eines möglichen Kandidaten im Zeitbereich mit ausgewerteten Größen, in der ein entsprechender Ausschnitt des Audio-Signals anhand der Energie in Dezibel dB im Verlauf der Zeit in Sekunden s zu sehen ist.

[0023] Ein erfindungsgemäßes Verfahren zum Erkennen eines akustischen Ereignisses in einem Audio-Signals S, und zwar in diesem Beispiel zum Erkennen einer Klappe, weist zwei Stufen A, B auf. In der ersten Stufe A werden mögliche Kandidaten X ausgewählt und in der zweiten Stufe B wird jedem der möglichen Kandidaten X ein Konfidenzwert W zugeordnet. Das Audio-Signal ist z. B. eine wav-Datei, die durch ein das erfindungsgemäße Verfahren durchführendes Programm bearbeitet wird.

[0024] Zur Auswahl der möglichen Kandidaten X weist die erste Stufe A des erfindungsgemäßen Verfahrens folgende in Figur 1 dargestellte Schritte auf:

Anwendung 110 eines ersten Hoch-Pass-Filters auf das Audio-Signal S, Berechnung 120 einer Energieeinhüllenden im Zeitbereich aus dem gefilterten Audio-Signal S, Berechnung 130 einer Ableitung aus der Energieeinhüllenden und Bestimmung 140 von möglichen Kandidaten aus Ereignisse, deren Maximalwert der Ableitung über einem vorbestimmten Schwellenwert liegt. Die Ableitung ist ein Maß für den Energieanstieg.

[0025] Der erste Hoch-Pass-Filter ist mit einer sehr flachen Flanke ausgeführt, d.h. er weist ein breites Übergangsband von z.B. Frequenzen zwischen 2000 und 3000 Hz auf. Dabei werden Frequenzen umso besser durchgelassen je höher sie sind, so dass höhere Frequenzen stärker gewichtet sind. Vorteil dieses Hoch-Pass-Filters ist auch, dass ein Filter mit einer solch flachen Flanke mit einer geringen Filterordnung und damit mit einer geringen Rechenkomplexität erreicht werden kann.

[0026] In dem Audio-Signal S ist bei einem möglichen Kandidat X der Maximalwert der Ableitung oberhalb eines bestimmten Schwellenwertes. Der Schwellenwert ist vom zu erkennenden Ereignis abhängig gewählt. In diesem Beispiel zum Erkennen von Klappen kann der Schwellenwert z. B. 18 dB betragen.

[0027] Da ein Klappenereignis innerhalb einer Genauigkeit eines viertel Bildes liegen sollte, d. h. im Bereich von 10 ms bei 25 Bildern pro Sekunde, wird zur Berechnung der Energieeinhüllenden ein Rechteckfenster F von 5 ms verwendet. Dieses Vorgehen entspricht einem Tief-Pass-Filter und ist geeignet, Rauschen zu unterdrücken.

[0028] In Figur 2 ist ein in der ersten Stufe A gefundener möglicher Kandidat X dargestellt. Das Rechteckfenster F ist in Figur 2 eingezeichnet.

[0029] Die zweiten Stufe B weist für jeden möglichen Kandidaten X folgende in Figur 1 dargestellte Schritte auf:

Anwendung 150 eines zweiten Hoch-Pass-Filters auf das Audio-Signal S, Berechnung 160 einer zweiten Energieeinhüllenden E im Zeitbereich aus dem gefilterten Audio-Signal S,
Auswertung 170 einer oder mehrere Größen mit Hilfe der Berechnung 160 der Energieeinhüllenden E und mit Hilfe

einer Bestimmung 180 eines Geräuschbereichs des gefilterten Audio-Signals S, und
Zuordnung 190 eines gemeinsamen Konfidenzwertes W mit Hilfe einer Bewertung 200 der Größen.

[0030] Der Konfidenzwert W ist ein Maß für die Wahrscheinlichkeit, dass es sich um das gesuchte Ereignis handelt. Als relatives Maß verglichen mit Konfidenzwerten W weiterer möglicher Kandidaten X eines Audio-Signals ermöglicht der Konfidenzwert W ein schnelles Auffinden des richtigen Kandidaten.

[0031] Die Auswertung 180 der Größen eines möglichen Kandidaten X erfolgt zusätzlich mit Hilfe des in der ersten Stufe A ermittelten Maximalwertes der Ableitung, d.h. des Energieanstiegs.

[0032] Der zweite Hoch-Pass-Filter, der auf das ursprüngliche Audio-Signal S angewandt wird, hat eine Grenzfrequenz von z.B. 200 Hz. Er ist eingesetzt, um Geräusche mit einer tiefen Frequenz, wie z. B. ein 50 Hz - oder 60 Hz - Brummen oder mechanische Geräusche einer laufenden Kamera zu unterdrücken.

[0033] Die Bestimmung 180 des Geräuschbereichs umfasst eine Bestimmung eines Grundrauschens G und/oder eines Aufnahmepegel A des Audio-Signals S. Bei der Bestimmung des Grundrauschens G und bei der Bestimmung des Aufnahmepegels A wird die in der zweiten Stufe B berechnete Energieeinhüllende E verwendet, wobei ein Histogramm von Werten der Energieeinhüllenden E erstellt wird. Als Aufnahmepegel A wird beispielsweise der Wert definiert, der nur von 1 % der Werte überschritten wird, und als Grundrauschen der Wert definiert, der von 5% der Werte nicht überschritten wird. Ausreißer mit sehr geringer Energie, z.B. durch ein Einschalten eines Mikrofons, werden bei diesem Verfahren nicht berücksichtigt. Außerdem ist der Aufnahmepegel A aus längeren Signalabschnitten zu ermitteln als das Grundrauschen G.

[0034] In der zweiten Stufe B erfolgt für jeden möglichen Kandidaten die Auswertung 170 einer oder mehrerer der folgenden Größen:

- Energieanstieg, d.h. der Maximalwert der Ableitung ,
- Höhe und Position des gemessenen Maximums M,
- Steigung und Abweichung von einer an den Energieabfall der Einhüllenden angepassten Kurve K,
- Differenz zwischen einem gemessenen Maximum M und einem aus der Kurve K vorausgesagten Maximum,
- Dauer T des möglichen Kandidaten X,
- Dauer T_v einer Stilleperiode vor dem möglichen Kandidaten X und Dauer T_n einer Stilleperiode nach dem möglichen Kandidaten X, und
- Zeitpunkt t_x des Auftreten des möglichen Kandidaten X.

[0035] Der Energieanstieg ist die einzige Größe, die in der ersten Stufe A ermittelt wird, und die aus der Energieeinhüllenden des durch den ersten Hoch-Pass-Filter gefilterten Audio-Signals S berechnet wird. Alle anderen Größen werden von der Energieeinhüllenden E des durch den zweiten, nur tiefe Frequenzen abschneidenden Hoch-Pass-Filter gefilterten Audio-Signals S abgeleitet, die in der zweiten Stufe B ermittelt wird.

[0036] Bei der Auswertung des gemessenen Maximums M wird für seine Höhe des Maximums M die Differenz zwischen dem gemessenen Maximum und dem Aufnahmepegel A ermittelt. Außerdem wird seine Position festgestellt. Ein gefundenes Maximum wird durch ein früheres lokales Maximum ersetzt, wenn es vermutlich durch Reflektionen erzeugt ist. Dazu wird das Maximum in zwei unterschiedlichen Zeitintervallen bestimmt, in einem kürzeren und in einem längeren. Das Maximum im längeren Zeitintervall muss signifikant höher sein, um als reales Maximum akzeptiert zu werden.

[0037] Es wird die Steigung und Abweichung von einer an den Energieabfall der Einhüllenden angepassten Kurve K ausgewertet. Diese Auswertung berücksichtigt, dass der Energieabfall des Klappenereignisses durch die Reflektionen im Raum, d.h. an den Wänden, am Boden und an der Decke, exponentiell abfällt. Die Anpassung der Kurve erfolgt in einer logarithmischen Skalierung, so dass eine einfache Anpassung an einen linearen Abfall erfolgt. Außerdem ermöglicht diese Anpassung, die Qualität der Anpassung durch die mittlere quadratische Abweichung festzustellen.

[0038] Ein exponentieller Energieabfall stellt sich bei der Energieeinhüllenden E normalerweise erst im hinteren Verlauf durch spätere diffuse Reflektionen im so genannten Nachhall ein. Im Anfangsbereich wirken sich eher diskrete Reflektionen auf den Abfall aus. Daher wird die Kurvenanpassung auf den hinteren Teil des akustischen Ereignisses beschränkt. Bei der Kurvenanpassung werden Messwerte in Abhängigkeit zu ihrem Abstand zum Grundrauschen G gewichtet, da Werte niedriger Energie, d.h. nahe am Grundrauschen G, stärker von Hintergrundgeräuschen beeinflusst werden. Die Kurvenanpassung wird gering bewertet, wenn das Audio-Signal S vermutlich außen aufgenommen wurde, d.h. wenn es kurz ist und nur diskrete Reflektionen und kaum Nachhall vorhanden sind. Dies erfolgt, indem die Dauer des möglichen Kandidaten X und eine sigmoidale Gewichtungsfunktion benutzt werden.

[0039] Der Energieabfall kann durch simultane Hintergrundgeräusche oder andere Vordergrundgeräusche unterbrochen sein. In diesem Fall wird die Kurvenanpassung nur bis zu dieser Unterbrechung durchgeführt. Zur Erkennung einer Unterbrechung wird auf die Energieeinhüllende E ein zusätzlicher Tief-Pass-Filter angewandt. Eine Unterbrechung des Energieabfalls wird festgestellt, wenn diese gefilterte Energieeinhüllende wieder ansteigt bevor die originale Energieeinhüllende E einen unteren Stilleschwellenwert S_1 erreicht. Bei Feststellung einer Unterbrechung des Energieabfalls

wird der Konfidenzwert W des möglichen Kandidaten X direkt oder indirekt in Abhängigkeit vom Abstand der Unterbrechung zu einem unteren Stilleschwellenwert S_1 verringert.

[0040] Die Differenz zwischen einem gemessenen Maximum M und einem aus der Kurve K vorausgesagten Maximum wird in einer logarithmischen Skalierung ermittelt. Es handelt sich daher um eine relative Differenz.

[0041] Die Dauer T des möglichen Kandidaten X , d.h. des akustischen Ereignisses, wird aus der Zeitspanne ermittelt, in der sich die Energie, d.h. die Energieeinhüllende E , über dem unteren Stilleschwellenwert S_1 befindet.

[0042] Die Dauer T_v einer Stilleperiode vor dem akustischen Ereignis, d.h. vor dem möglichen Kandidaten X , und die Dauer T_n einer Stilleperiode nach dem möglichen Kandidaten X sind Zeitspannen, die die Energieeinhüllende E benötigt, um über einen oberen Stilleschwellenwert S_2 zu kommen, nachdem sie unter den unteren Stilleschwellenwert S_1 gefallen ist. Diese Hysterese verhindert, dass leise Geräusche als Ende einer Stilleperiode erkannt werden. Bei einer richtigen Klappe sind die Stilleperioden T_v und T_n weder zu lang noch zu kurz. Wenn die Bewegung zum Schließen selbst Geräusche verursacht, gibt es möglicherweise keine Stilleperiode T_v vor der Klappe. Dies wird bei der Auswertung berücksichtigt. Für Außenaufnahmen werden bei der Auswertung der Stilleperioden T_v und T_n , soweit möglich, Echos vernachlässigt.

[0043] Bei der Auswertung des Zeitpunktes t_x des Auftretens des möglichen Kandidaten wird berücksichtigt, dass sich ein möglicher Kandidat, nämlich eine Klappe, typischerweise am Anfang oder am Ende einer Aufnahme befindet.

[0044] Die zweite Stufe B umfasst für die Bewertung 200 der oben beschriebenen, ausgewerteten Größe folgende Schritte für jede Größe: Bestimmung eines Wahrscheinlichkeitsverhältnisses v und/oder eines Gewichtungsfaktor w .

[0045] Bei der Zuordnung 190 eines gemeinsamen Konfidenzwertes W zu einem möglichen Kandidaten werden die Wahrscheinlichkeitsverhältnisse v und/oder der Gewichtungsfaktoren w der ausgewerteten Größen zusammengefasst. Dies erfolgt durch eine Addition der Logarithmen der durch die Gewichtungsfaktoren w gewichteten Wahrscheinlichkeitsverhältnisse v der ausgewählten Größen. Die Gewichtungsfaktoren w der ausgewerteten Größen werden jeweils aus Korrelationskoeffizienten k für paarweise Korrelationen der ausgewerteten Größen berechnet werden.

[0046] Insbesondere wird bei N ausgewerteten Größen der Gewichtungsfaktor w_i einer Größe i aus den Korrelationskoeffizienten k_{ij} für die N paarweisen Korrelationen folgendermaßen berechnet:

$$W_i = \frac{1}{\sum_j |k_{ij}|^m}, \quad j = 1 \text{ bis } N$$

[0047] Der Korrelationskoeffizienten k_{ij} ist ein Maß für die Korrelation zwischen der i -ten und j -ten Größe und wird aus empirischen Daten ermittelt. Bei der Berechnung der Korrelationskoeffizienten k_{ij} werden Ausreißer, die eine 3σ -Grenze überschreiten, unterdrückt. Der Exponent m bestimmt, wie stark die Korrelation in Betracht gezogen wird. Je größer der Exponent m , desto geringer wird der Einfluß einer möglichen Korrelation berücksichtigt. Er sollte höher gewählt werden, wenn nur wenige Daten zum Abschätzen der Korrelationskoeffizienten vorhanden sind.

[0048] In einer alternativen Ausführungsform der Erfindung werden bei der Bestimmung der Wahrscheinlichkeitsverhältnisse v eine oder mehrere Zusatzinformationen über das akustische Ereignis berücksichtigt. Derartige Zusatzinformationen sind z.B. folgende Informationen über das Audio-Signal S :

Separate Aufnahmen mit Anfangsklappen oder Endklappen,
Soloklappen, oder
Innenaufnahmen oder Außenaufnahmen.

[0049] In einer weiteren alternativen Ausführungen der Erfindung umfasst die zweite Stufe B für jeden möglichen Kandidaten X alternativ oder zusätzlich folgenden Schritt:

Spracherkennung eines auf das akustische Ereignis hinweisenden Textes.

Patentansprüche

1. Verfahren zum Erkennen eines akustischen Ereignisses eines Audio-Signals (S), bei dem in einer ersten Stufe (A) mögliche Kandidaten (X) ausgewählt werden und

in einer zweiten Stufe (B) jedem der möglichen Kandidaten (X) ein Konfidenzwert (W) zugeordnet wird.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** die erste Stufe (A) folgende Schritte aufweist:

Anwendung (110) eines ersten Hoch-Pass-Filters auf das Audio-Signal (S), Berechnung (120) einer Energieeinhüllenden im Zeitbereich aus dem gefilterten Audio-Signal (S),
Berechnung (130) einer Ableitung aus der Energieeinhüllenden und
Bestimmung (140) von möglichen Kandidaten aus Ereignisse, deren Maximalwert der Ableitung über einem vorbestimmten Schwellenwert liegt.

3. Verfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** die zweiten Stufe (B) für jeden möglichen Kandidaten (X) folgende Schritte aufweist:

Auswertung (170) einer oder mehrere Größen und
Zuordnung (190) eines gemeinsamen Konfidenzwertes (W) mit Hilfe einer Bewertung (200) der Größen.

4. Verfahren nach Anspruch 3, **dadurch gekennzeichnet, dass** in der zweiten Stufe (B) für jeden möglichen Kandidaten (X) die Auswertung (170) einer oder mehrerer der folgenden Größen erfolgt:

- Energieanstieg (E), d.h. des Maximalwertes der Ableitung,
- Höhe und Position des gemessenen Maximums (M),
- Steigung und Abweichung von einer an den Energieabfall der Einhüllenden (E) angepassten Kurve (K),
- Differenz zwischen einem gemessenen Maximum (M) und einem aus der Kurve (K) vorausgesagten Maximum,
- Dauer (T) des möglichen Kandidaten (X),
- Dauer (T_v) einer Stilleperiode vor dem möglichen Kandidaten und Dauer (T_n) einer Stilleperiode nach dem möglichen Kandidaten (X), und
- Zeitpunkt (t_x) des Auftretens des möglichen Kandidaten (X).

5. Verfahren nach Anspruch 4, **dadurch gekennzeichnet, dass** die zweite Stufe (B) für jeden möglichen Kandidaten (X) für die Auswertung (170) der Größen folgende Schritte aufweist:

Anwendung (150) eines zweiten Hoch-Pass-Filters auf das Audio-Signal (S) und Berechnung (160) einer Energieeinhüllenden (E) im Zeitbereich aus dem gefilterten Audio-Signal (S).

6. Verfahren nach Anspruch 4 oder 5, **dadurch gekennzeichnet, dass** die zweite Stufe (B) für jeden möglichen Kandidaten (X) für die Auswertung (170) der Größen folgenden Schritt aufweist:

Bestimmung (180) eines Geräuschbereichs des Audio-Signals (S).

7. Verfahren nach Anspruch 6, **dadurch gekennzeichnet, dass** die Bestimmung (180) des Geräuschbereichs eine Bestimmung eines Grundrauschens (G) und/oder eines Aufnahmepegels (A) umfasst.

8. Verfahren nach Anspruch 6 oder 7, **dadurch gekennzeichnet, dass** bei der Bestimmung (180) des Geräuschbereichs, die in der zweiten Stufe (B) berechnete Energieeinhüllende (E) verwendet wird.

9. Verfahren nach einem der Ansprüche 3 bis 8, **dadurch gekennzeichnet, dass** die zweite Stufe (B) für jeden möglichen Kandidaten (X) für die Bewertung (200) einer oder mehrerer der ausgewerteten Größe jeweils folgende Schritte aufweist:

Bestimmung eines Wahrscheinlichkeitsverhältnisses (v) und/oder eines Gewichtungsfaktor (w).

10. Verfahren nach Anspruch 9, **dadurch gekennzeichnet, dass** bei der Zuordnung (190) eines gemeinsamen Konfidenzwertes (W) die Wahrscheinlichkeitsverhältnisse (v) und/oder der Gewichtungsfaktoren (w) der ausgewerteten Größen zusammengefasst werden.

EP 2 031 581 A1

11. Verfahren nach Anspruch 10, **dadurch gekennzeichnet, dass** bei der Zuordnung (190) eines gemeinsamen Konfidenzwertes (W) eine Addition der Logarithmen der durch die Gewichtungsfaktoren (w) gewichteten Wahrscheinlichkeitsverhältnisse (v) der ausgewählten Größen erfolgt.

5 12. Verfahren nach einem der Ansprüche 9 bis 11, **dadurch gekennzeichnet, dass** die Gewichtungsfaktoren (w) einer oder mehrerer der ausgewerteten Größen jeweils aus Korrelationskoeffizienten (k) für paarweise Korrelationen der ausgewerteten Größen berechnet werden.

10 13. Verfahren nach einem der Ansprüche 9 bis 12, **dadurch gekennzeichnet, dass** bei der Bestimmung der Wahrscheinlichkeitsverhältnisse (v) eine oder mehrere Zusatzinformationen über das akustische Ereignis berücksichtigt werden.

15 14. Verfahren nach einem der Ansprüche 1 bis 13, **dadurch gekennzeichnet, dass** die zweite Stufe (B) für jeden möglichen Kandidaten (X) alternativ oder zusätzlich folgenden Schritt aufweist:

Spracherkennung eines auf das akustische Ereignis hinweisenden Textes.

20 15. Verfahren nach einem der Ansprüche 1 bis 14, das zum Erkennen von Klappen bei der Synchronisation des Audio-Signals (S) mit einem entsprechenden Video-Signal verwendet wird.

25

30

35

40

45

50

55

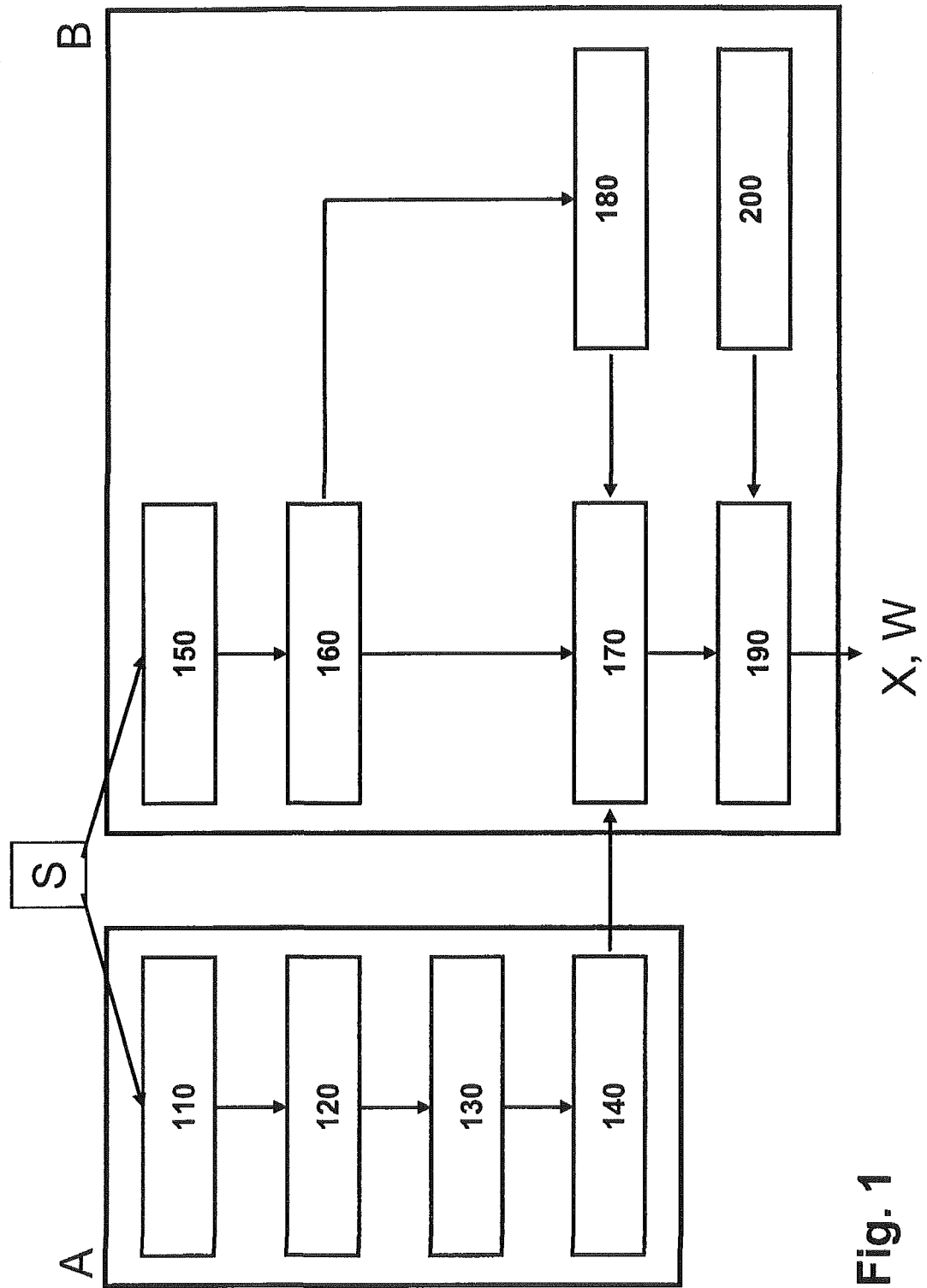


Fig. 1

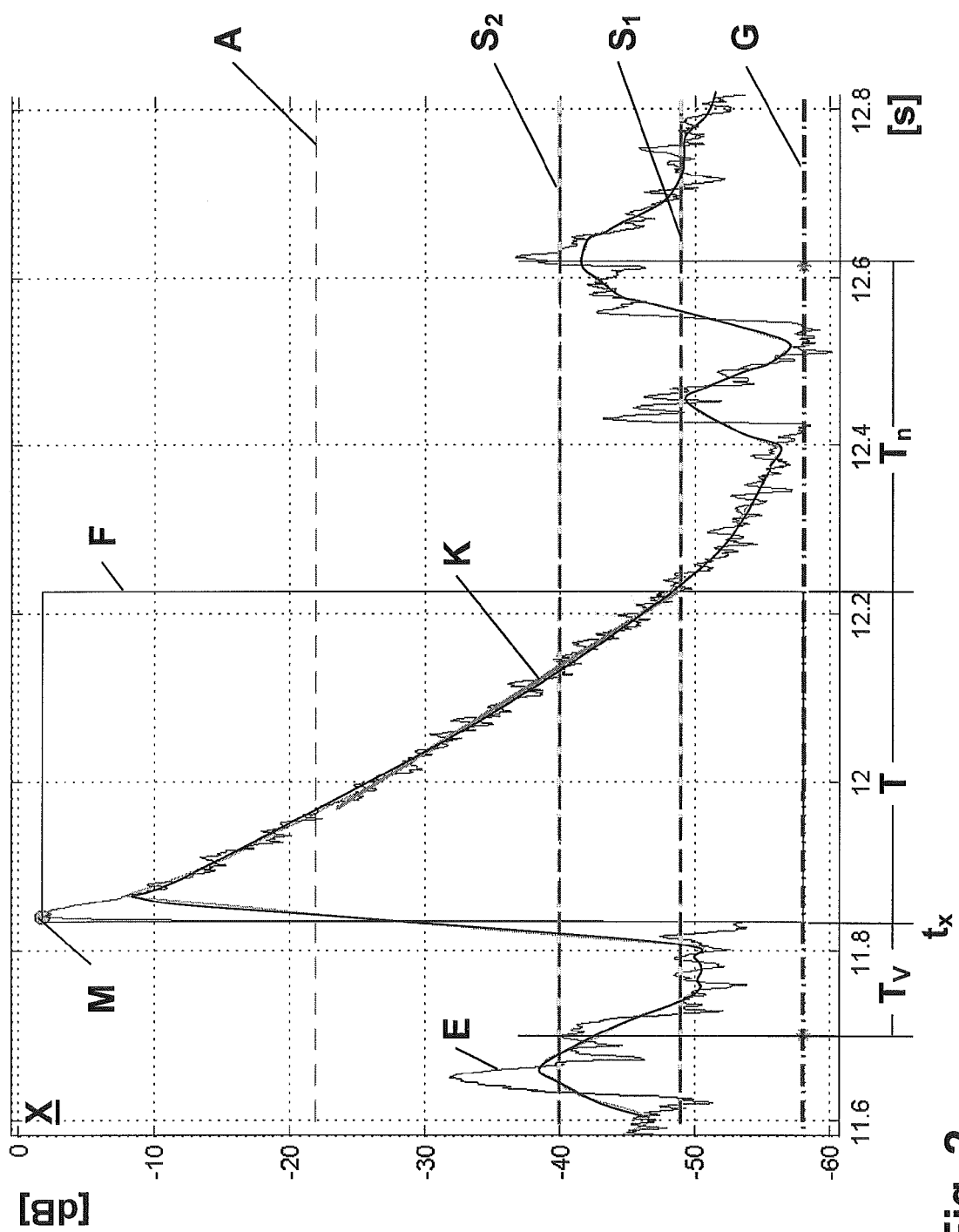


Fig. 2



Europäisches
Patentamt

EUROPÄISCHER RECHERCHENBERICHT

Nummer der Anmeldung
EP 07 11 5473

EINSCHLÄGIGE DOKUMENTE			
Kategorie	Kennzeichnung des Dokuments mit Angabe, soweit erforderlich, der maßgeblichen Teile	Betrifft Anspruch	KLASSIFIKATION DER ANMELDUNG (IPC)
X	CHI HANG WONG ET AL: "Automatic lyrics alignment for Cantonese popular music" MULTIMEDIA SYSTEMS, SPRINGER-VERLAG, BE, Bd. 12, Nr. 4-5, 12. September 2006 (2006-09-12), Seiten 307-323, XP019492688 ISSN: 1432-1882 * Seite 311, rechte Spalte, Zeile 19 - Seite 312, rechte Spalte, Zeile 3 * * Abschnitte 4.3,5 * * Abbildungen 4,6 *	1-8,14, 15	INV. G10L11/00
A	-----	9-13	
X	US 6 787 689 B1 (CHEN FANG-CHU [TW]) 7. September 2004 (2004-09-07) * Spalte 5, Zeilen 33-44 * * Spalte 5, Zeile 66 - Spalte 6, Zeile 5 * * Spalte 6, Zeilen 32-51 * * Spalte 7, Zeilen 31-61 * * Abbildungen 1,3 *	1,3-8, 14,15	
X	KLAPURI A: "Sound onset detection by applying psychoacoustic knowledge" ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1999. PROCEEDINGS., 1999 IEEE INTERNATIONAL CONFERENCE ON PHOENIX, AZ, USA 15-19 MARCH 1999, PISCATAWAY, NJ, USA, IEEE, US, Bd. 6, 15. März 1999 (1999-03-15), Seiten 3089-3092, XP010328057 ISBN: 0-7803-5041-3 * Abschnitte 3,4 * * Abbildungen 1b,3 *	1,3-8, 14,15	RECHERCHIERTE SACHGEBIETE (IPC) G10L G03B
Der vorliegende Recherchenbericht wurde für alle Patentansprüche erstellt			
Recherchenort München		Abschlußdatum der Recherche 23. Oktober 2007	Prüfer Geißler, Christian
KATEGORIE DER GENANNTEN DOKUMENTE X : von besonderer Bedeutung allein betrachtet Y : von besonderer Bedeutung in Verbindung mit einer anderen Veröffentlichung derselben Kategorie A : technologischer Hintergrund O : nichtschriftliche Offenbarung P : Zwischenliteratur		T : der Erfindung zugrunde liegende Theorien oder Grundsätze E : älteres Patentdokument, das jedoch erst am oder nach dem Anmeldedatum veröffentlicht worden ist D : in der Anmeldung angeführtes Dokument L : aus anderen Gründen angeführtes Dokument & : Mitglied der gleichen Patentfamilie, übereinstimmendes Dokument	

EPO FORM 1503 03.82 (P04C03)



Europäisches
Patentamt

EUROPÄISCHER RECHERCHENBERICHT

Nummer der Anmeldung
EP 07 11 5473

EINSCHLÄGIGE DOKUMENTE			
Kategorie	Kennzeichnung des Dokuments mit Angabe, soweit erforderlich, der maßgeblichen Teile	Betrifft Anspruch	KLASSIFIKATION DER ANMELDUNG (IPC)
X	GAINZA M ET AL: "Onset based audio segmentation for the irish tin whistle" SIGNAL PROCESSING, 2004. PROCEEDINGS. ICSP '04. 2004 7TH INTERNATIONAL CONFERENCE ON BEIJING, CHINA AUG. 31 - SEPT 4, 2004, PISCATAWAY, NJ, USA, IEEE, 31. August 2004 (2004-08-31), Seiten 594-597, XP010809694 ISBN: 0-7803-8406-7 * Seite 595, rechte Spalte, Zeilen 20-29 * * Seite 596, linke Spalte, Zeilen 7-27 * * Seite 596, rechte Spalte, Zeilen 4-22 * -----	1,3-8, 14,15	
X	WO 2006/129274 A (KONINKL PHILIPS ELECTRONICS NV [NL]; SKOWRONEK JANTO [NL]; MCKINNEY MA) 7. Dezember 2006 (2006-12-07) * Seite 6, Zeile 25 - Seite 7, Zeile 3 * * Abbildungen 2,4 * -----	1,3-8, 14,15	
			RECHERCHIERTE SACHGEBIETE (IPC)
Der vorliegende Recherchenbericht wurde für alle Patentansprüche erstellt			
Recherchenort München		Abschlußdatum der Recherche 23. Oktober 2007	Prüfer Geißler, Christian
KATEGORIE DER GENANNTEN DOKUMENTE X : von besonderer Bedeutung allein betrachtet Y : von besonderer Bedeutung in Verbindung mit einer anderen Veröffentlichung derselben Kategorie A : technologischer Hintergrund O : mündliche Offenbarung P : Zwischenliteratur		T : der Erfindung zugrunde liegende Theorien oder Grundsätze E : älteres Patentdokument, das jedoch erst am oder nach dem Anmeldedatum veröffentlicht worden ist D : in der Anmeldung angeführtes Dokument L : aus anderen Gründen angeführtes Dokument & : Mitglied der gleichen Patentfamilie, übereinstimmendes Dokument	

3
EPO FORM 1503 03 82 (P04C03)

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

EP 07 11 5473

In diesem Anhang sind die Mitglieder der Patentfamilien der im obengenannten europäischen Recherchenbericht angeführten Patentdokumente angegeben.

Die Angaben über die Familienmitglieder entsprechen dem Stand der Datei des Europäischen Patentamts am
 Diese Angaben dienen nur zur Unterrichtung und erfolgen ohne Gewähr.

23-10-2007

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 6787689 B1	07-09-2004	KEINE	
WO 2006129274 A	07-12-2006	KEINE	

EPO FORM P0461

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

IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

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

In der Beschreibung aufgeführte Patentdokumente

- EP 1465192 A1 [0003]
- US 5884260 A [0004]
- US 20050199064 A1 [0004]