



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) DE 600 23 517 T2 2006.06.01

(12)

## Übersetzung der europäischen Patentschrift

(97) EP 1 210 711 B1

(51) Int Cl.<sup>8</sup>: G10L 17/00 (2006.01)

(21) Deutsches Aktenzeichen: 600 23 517.3

(86) PCT-Aktenzeichen: PCT/US00/23754

(96) Europäisches Aktenzeichen: 00 963 280.3

(87) PCT-Veröffentlichungs-Nr.: WO 01/016937

(86) PCT-Anmeldetag: 29.08.2000

(87) Veröffentlichungstag

der PCT-Anmeldung: 08.03.2001

(97) Erstveröffentlichung durch das EPA: 05.06.2002

(97) Veröffentlichungstag

der Patenterteilung beim EPA: 26.10.2005

(47) Veröffentlichungstag im Patentblatt: 01.06.2006

(30) Unionspriorität:

385975 30.08.1999 US

(74) Vertreter:

Grünecker, Kinkeldey, Stockmair &  
Schwanhäusser, 80538 München

(73) Patentinhaber:

Harman Becker Automotive  
Systems-Wavemakers, Inc., Vancouver, British  
Columbia, CA

(84) Benannte Vertragsstaaten:

AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT,  
LI, LU, MC, NL, PT, SE

(72) Erfinder:

ZAKARAUSKAS, Pierre, Encinitas, US

(54) Bezeichnung: KLASSIFIZIERUNG VON SCHALLQUELLEN

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelebt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

**Beschreibung****TECHNISCHES GEBIET**

**[0001]** Diese Erfindung betrifft Systeme und Verfahren für die automatische Klassifikation von akustischen (Schall-) Quellen, einschließlich einer text-unabhängigen Sprecheridentifikation.

**HINTERGRUND**

**[0002]** Es gibt einige Forschungsgebiete, die die Klassifikation akustischer Signale studieren. Jedes Forschungsgebiet hat seine eigenen Ansätze zur Klassifikation akustischer Signale, mit einigen Überlappung zwischen ihnen. Zur Zeit sind die Hauptanwendungen für die automatische Klassifikation von Schallquellen: Sprecherverifikation; Sprecheridentifikation; passive Sonarklassifikation und das Überwachen oder die Diagnose von Maschinengeräuschen.

**[0003]** Die Sprecherverifikation zielt darauf, zu verifizieren, dass ein gegebener Sprecher in der Tat ist, wer er oder sie behauptet, zu sein. Bei den meisten Sprecherverifikationssystemen arbeitet ein Sprecher mit, indem er ein Schlüsselwort sagt, und das System vergleicht die Art, in der das Schlüsselwort von dem mutmaßlichen Sprecher gesagt wurde mit Trainingsvorlagen der gleichen Schlüsselwörter. Wenn die Passung schlecht ist, wird der Sprecher zurückgewiesen oder der Service (z. B. der Computer- oder Gebäudezugang) verweigert. Ein Nachteil solcher Verfahren besteht darin, dass das gleiche Schlüsselwort zu der Prüfzeit wie zu der Trainingszeit verwendet werden muss, wodurch die Anwendung solcher Verfahren auf die Zugriffssteuerung eingeschränkt wird. Dieses Verfahren könnte zum Beispiel nicht verwendet werden, um die Sprecher in einer Hin-und-Zurück-Konversation zu bezeichnen.

**[0004]** Die Sprecheridentifikation zielt darauf, festzustellen, welcher von einem Satz von Stimmen am besten zu einer gegebenen Testäußerung passt. Die text-unabhängige Sprecheridentifikation versucht, eine solche Ermittlung ohne den Gebrauch von bestimmten Schlüsselwörtern auszuführen.

**[0005]** Die passive Sonarklassifikation schließt die Identifikation eines Schiffs entsprechend dem Schall, den es unter Wasser abstrahlt, ein. Das Überwachen und die Diagnose von Maschinengeräuschen schließt das Bestimmen des Zustands eines Maschinenteils über den Schall, den es abgibt, ein.

**[0006]** In sämtlichen der oben genannten Anwendungen, wird zuerst ein Modell jeder Schallquelle erhalten, indem man ein System mit einem Satz von Beispieldämpfen von jeder Quelle trainiert. Eine Testvorlage wird dann mit den gespeicherten Modellen verglichen, um eine Schallquellenkategorie für die Testvorlage festzustellen. Bekannte Verfahren erfordern verhältnismäßig lange Trainingszeiten und viele Testvorlagen, wodurch solche Verfahren in vielen Fällen ungeeignet sind. Des Weiteren weisen solche Verfahren die Tendenz auf, eine große Menge an Speicher und große Computerressourcen zu erfordern. Schließlich sind diese Verfahren häufig nicht gegenüber der Gegenwart von Geräuschen in dem Testsignal Fehler unanfällig, was ihre Verwendung für viele Aufgaben verhindert. ("Signal" bedeutet ein Signal, das von Interesse ist; Hintergrund- und Ablenkungstone werden als "Geräusche" bezeichnet).

**[0007]** Der Erfinder hat festgestellt, dass es wünschenswert sein würde, in der Lage zu sein, ein akustisches Signal zu klassifizieren, selbst wenn einige Teile der Spektren durch Geräusche verdeckt werden, und dass ein minimaler Trainings- und Testaufwand erforderlich ist. Die vorliegende Erfindung stellt ein System und ein Verfahren für die Klassifikation akustischer Signale zur Verfügung, das die Einschränkungen des Standes der Technik vermeidet.

**ZUSAMMENFASSUNG**

**[0008]** Die Erfindung schließt ein Verfahren, eine Vorrichtung und ein Computerprogramm zur Klassifizierung einer Schallquelle ein. Die Erfindung vergleicht den akustischen Eingang mit einer Anzahl von Signalmustern, einem pro Quellenklasse, und erzeugt eine Bewertung für jedes Signalmuster. Die Schallquelle wird als derselben Klasse zugehörig erklärt, wie die des Modells mit der besten Bewertung, wenn diese Bewertung hinreichend hoch ist. In der bevorzugten Ausführungsform wird die Klassifikation durch den Gebrauch von einem Signalmuster erreicht, das durch Lernen vergrößert wird. Das Eingangssignal kann menschliche Rede darstellen, in welchem Fall es das Ziel sein würde, den Sprecher in einer text-unabhängigen Weise zu identifizieren. Jedoch sollte es erkannt werden, dass die Erfindung verwendet werden kann, um irgendeine Art von akustischen Livedaten oder aufgezeichneten akustischen Daten, wie Musikinstrumente, Vögel, Motor- oder Maschi-

nengeräusch oder menschliches Singen zu klassifizieren. Die bevorzugte Ausführungsform der Erfindung klassifiziert Eingangssignale wie folgt. Ein Eingangssignal wird in binäre Daten digitalisiert, die in eine Zeit-Frequenzdarstellung (Spektrogramm) umgewandelt werden. Das Hintergrundgeräusch wird bestimmt, und es isoliert ein Signaldetektor Perioden, die ein Signal enthalten. Perioden ohne Signalinhalt sind in der Geräuschschätzung eingeschlossen. Das Spektrogramm des Eingangssignals wird umskaliert und mit Spektrogrammen für eine Anzahl von Vorlagen, die ein Signalmodell definieren, verglichen, wobei jedes Signalmodell eine Quellenklasse darstellt. Die durchschnittliche Abweichung zwischen den gemessenen Spektrogrammen und den Spektrogrammen jedes Signalmodells wird errechnet. Das Signalmodell mit der niedrigsten Abweichung wird ausgewählt. Wenn die durchschnittliche Abweichung des vorgewählten Signalmodells hinreichend klein ist, wird die Quelle als der entsprechenden Klasse zugehörig erklärt. Wenn nicht, wird erklärt, dass die Quelle von unbekannter Art ist.

**[0009]** Die Menge der Signalmodelle wird mit Signaldaten trainiert, indem man Vorlagen von den Spektrogrammen der Eingangssignale herstellt, wenn solche Spektrogramme zu den Spektrogrammen der vorhandenen Vorlagen signifikant unterschiedlich sind. Wenn eine vorhandene Vorlage gefunden wird, die dem Eingangssignal-Spektrogramm ähnelt, wird diese Vorlage mit dem Eingangssignal-Spektrogramm auf eine solche Weise gemittelt, dass die resultierende Vorlage den Durchschnitt von sämtlichen Spektren darstellt, die in der Vergangenheit zu der Vorlage passten.

**[0010]** Die Erfindung hat die folgenden Vorteile: Sie ist in der Lage, eine akustische Signalquelle zu klassifizieren: unabhängig von dem Ton, den die Quelle zu der Zeit des Samplings abstrahlt; unabhängig von den Schallpegeln; und selbst wenn einige Teile der Spektren des akustischen Signals durch Geräusch überdeckt werden. Die Erfindung erfordert auch verhältnismäßig wenig Training, Testdaten und Computerressourcen.

**[0011]** Die Einzelheiten von einer oder mehr Ausführungsformen der Erfindung werden in den begleitenden Zeichnungen und in der untenstehenden Beschreibung dargelegt. Andere Merkmale, Gegenstände und Vorteile der Erfindung werden aus der Beschreibung und den Zeichnungen und aus den Ansprüchen offensichtlich.

#### BESCHREIBUNG DER ZEICHNUNGEN

**[0012]** [Fig. 1](#) ist ein Blockdiagramm eines programmierbaren Computersystems nach dem Stand der Technik, das für das Implementieren der Signalverbesserungstechnik der Erfindung geeignet ist.

**[0013]** [Fig. 2](#) ist ein Flussdiagramm, welches das grundlegende Verfahren der bevorzugten Ausführungsform der Erfindung zeigt.

**[0014]** [Fig. 3](#) ist ein Flussdiagramm, das einen bevorzugten Prozess für das Schätzen von Hintergrundgeräuschparametern und das Ermitteln des Vorhandenseins des Signals zeigt.

**[0015]** [Fig. 4](#) ist ein Flussdiagramm, welches das bevorzugte Verfahren zeigt, um das Vorhandensein der harmonisch aufeinander bezogenen Spitzen in einem Signalspektrum zu ermitteln.

**[0016]** [Fig. 5](#) ist ein Flussdiagramm, das ein bevorzugtes Verfahren für das Erzeugen und das Verwenden der Signalmodellvorlagen zeigt.

**[0017]** Gleiche Bezugszeichen und Bezeichnungen in den verschiedenen Zeichnungen bezeichnen gleiche Elemente.

#### AUSFÜHRLICHE BESCHREIBUNG

**[0018]** Durchgehend in dieser Beschreibung sollten die bevorzugten Ausführungsformen und gezeigten Beispiele exemplarisch anstatt als Einschränkungen der Erfindung angesehen werden.

#### Überblick über die Betriebsumgebung

**[0019]** [Fig. 1](#) zeigt ein Blockdiagramm eines typischen programmierbaren Verarbeitungssystems aus dem Stand der Technik, welches für das Implementieren des Klassifizierungssystems für akustische Signale der Erfindung verwendet werden kann. Ein akustisches Signal wird an einem Signalamformer-Mikrofon **10** empfangen, das eine entsprechende elektrische Signaldarstellung des akustischen Signals erzeugt. Das Signal von

dem Signalumformer mikrophon **10** wird dann vorzugsweise durch einen Verstärker **12** verstärkt, bevor es durch einen Analog-Digital-Wandler **14** digitalisiert wird. Die Ausgabe des Analog-Digital-Wandlers **14** wird einem Verarbeitungssystem zugeführt, das die Klassifikationstechniken der Erfindung verwendet. Das Verarbeitungssystem schließt vorzugsweise eine CPU **16**, RAM **20**, ROM **18** (der, wie ein Flash-ROM, beschreibbar sein kann) und eine optionale Speichereinrichtung **22**, wie eine magnetische Disk, ein, die über einen CPU-Bus, wie gezeigt, miteinander verbunden sind. Die Ausgabe des Klassifikationsprozesses kann zugunsten eines menschlichen Benutzers mittels eines Videoanzeigekontrollers **24** angezeigt werden, der eine Videoanzeige **26** betreibt, oder von dem System benutzt werden, um seine Antwort zur Identität der Schallquelle anzufertigen, oder sie kann benutzt werden, um eine externe Ausrüstung (z. B. Verriegelungseinheiten in einer Zugriffssteuerungsanwendung) zu betätigen.

#### Funktionsüberblick über das System

**[0020]** Das Folgende beschreibt die Funktionsbestandteile eines Klassifizierungssystems für akustische Signale. Ein erster Funktionsbestandteil der Erfindung ist ein Vorprozessor, der Eingangsdaten in eine Zeit-Frequenzdarstellung umwandelt. Die Muster der relativen Leistung in den unterschiedlichen Frequenzbändern und wie solche Muster sich kurzfristig ändern, werden durch das vorliegende System benutzt, um ein Eingangssignal zu klassifizieren.

**[0021]** Die zweiten und dritten Funktionsbestandteile der Erfindung sind eine Vorrichtung zum dynamischen Abschätzen eines Hintergrunds bzw. ein Signaldetektor, die im Tandem arbeiten. Ein Signaldetektor ist nützlich, um von ununterbrochenem Hintergrundgeräusch zu unterscheiden. Es ist wichtig, sicherzugehen, dass die Klassifikation nur auf einem Signal basiert und nicht durch Hintergrundgeräusch beeinflusst wird. Die dynamische Hintergrundgeräusch-Schätzfunktion ist zu einem Unterscheiden vorübergehender Töne von dem Hintergrundgeräusch und zum Schätzen nur des Hintergrundgeräusches in der Lage. In einer Ausführungsform arbeitet ein Leistungsdetektor in jedem von mehrfachen Frequenzbändern. Teile der Daten, die lediglich Geräusch enthalten, werden verwendet, um Mittelwert und Standardabweichung des Geräusches in Dezibel (dB) zu erzeugen. Wenn die Leistung den Mittelwert um mehr als eine spezifizierte Anzahl von Standardabweichungen in einem Frequenzband übersteigt, wird der entsprechende Zeitabschnitt als ein Signal enthaltend gekennzeichnet und wird nicht verwendet, um das Spektrum lediglich des Geräusches zu schätzen.

**[0022]** Der vierte Funktionsbestandteil der Erfindung ist ein harmonischer Detektor. In dem Fall harmonischer Töne, wird der harmonische Detektor auch benutzt, um eine Schätzung für die Grundfrequenz des Signals zur Verfügung zu stellen, die für die Klassifikation nützlich sein kann. Ein harmonischer Detektor ist ein nützlicher Filter zum Anwenden auf die Daten, da in vielen Fällen von Interesse (z. B. menschliche Stimme, Musik, Vokalsang, Motor und Maschinerie) das Signal eine harmonische Struktur aufweist. Eine bevorzugte Ausführungsform eines harmonischen Detektors ist unten beschrieben. Der harmonische Detektor zählt die Zahl harmonisch aufeinander bezogener Spitzen in dem Spektrum.

**[0023]** Der fünfte Funktionsbestandteil ist eine spektrale Umskalierungseinrichtung. Das Eingangssignal kann schwach oder stark, nah oder fern sein. Bevor gemessene Spektren mit Vorlagen von einem Modell verglichen werden, werden die gemessenen Spektren umskaliert, so dass der Zwischenmusterabstand nicht von der Gesamtlautstärke des Signals abhängt. In der bevorzugten Ausführungsform wird eine Gewichtung proportional zu dem Signal-zu-Rausch-Verhältnis (SNR) in Dezibel (dB) auf die Frequenzbänder während des Umskalierens angewendet. Die Gewichte werden nach unten und oben durch einen minimalen beziehungsweise maximalen Wert begrenzt. Die Spektren werden umskaliert, so dass der gewichtete Abstand zu jeder gespeicherten Vorlage minimiert wird.

**[0024]** Der sechste Funktionsbestandteil ist eine Musterabgleicheinrichtung. Die Musterabgleicheinrichtung vergleicht das Spektrogramm des Eingangssignals mit einer Menge von Signalmodellen, von denen jedes eine Klasse definiert. Jedes Signalmodell besteht aus einem Satz prototypischer Spektrogramme von kurzer Dauer ("Vorlagen"), die aus Signalen bekannter Identität erhalten werden. Das Signalmodell-Training wird vollendet, indem Spektrogramme gesammelt werden, die von den vorher gesammelten prototypischen Spektrogrammen signifikant verschieden sind. In der bevorzugten Ausführungsform ist das erste prototypische Spektrogramm das erste Eingangssignal-Spektrogramm, das ein Signal enthält, das signifikant über dem Geräuschpegel liegt. Für die folgenden Zeitepochen wird, wenn das Eingangssignal-Spektrogramm näher an irgendeinem vorhandenen prototypischen Spektrogramm als eine vorgewählte Abstandsschwelle liegt, dann dieses Eingangssignal-Spektrogramm mit dem nächsten prototypischen Spektrogramm gemittelt. Wenn das Eingangssignal-Spekrogramm weiter von irgendeinem prototypischen Spektrogramm als die vorgewählte Schwelle entfernt ist, wird dann das Eingangssignal-Spektrogramm dazu erklärt, ein neues prototypisches Spektrogramm

zu sein.

**[0025]** Der Abstand zwischen den Vorlagen und dem gemessenen Spektrogramm des Eingangssignals kann eine von verschiedenen geeigneten Metriken, wie der euklidische Abstand oder ein gewichteter euklidischer Abstand, sein. Für jede Signalmodellklasse wird die Vorlage mit dem kleinsten Abstand zu dem gemessenen Eingangssignal-Spektrogramm als das beste passende prototypische Spektrogramm für diese Klasse ausgewählt.

**[0026]** Der siebte Funktionsbestandteil ist eine Klassifikationseinrichtung. Eine Bewertung für jede Klasse wird für jedes Eingangssignal-sample angesammelt. Wenn genügend Daten von einer verwendbaren Anzahl von Eingangssignal-samples gesammelt worden sind, wird eine abschließende Klassifikationsentscheidung getroffen. Alternativ kann eine Entscheidung zu jeder möglichen gewünschten Zeit oder zu jedem Ereignis (zum Beispiel, wenn eine Redeperiode von einer signifikanten Ruheperiode gefolgt ist) erzwungen werden, und es kann die am besten passende Klasse zusammen mit der Bewertung an diesem Punkt wiedergegeben werden.

#### Überblick über das grundlegende Verfahren

**[0027]** [Fig. 2](#) ist ein Flussdiagramm der bevorzugten Verfahrensausführungsform der Erfindung. Das Verfahren, das in [Fig. 2](#) gezeigt wird, wird für das Verbessern eines eingehenden akustischen Signals verwendet, das aus einer Mehrzahl von Datensamples besteht, die als Ausgabe von dem Analog-Digital-Wandler **14** erzeugt werden, der in [Fig. 1](#) gezeigt wird. Das Verfahren beginnt an einem Anfangszustand (Schritt **202**). Der eingehende Datenstrom (z. B. eine vorher erzeugte Datei akustischer Daten oder ein digitalisiertes akustisches Live-signal) wird in einen Computerspeicher als Satz von Samples eingelesen (Schritt **204**). In der bevorzugten Ausführungsform würde die Erfindung normalerweise verwendet, um aus einem „sich bewegenden Fenster“ von Daten zu klassifizieren, die Teile eines ununterbrochenen akustischen Datenstromes darstellen, so dass der gesamte Datenstrom verarbeitet wird. Im Allgemeinen wird ein akustischer Datenstrom, der zu klassifizieren ist, als Reihe von Daten "puffern" von fester Länge unabhängig von der Dauer des ursprünglichen akustischen Datenstromes dargestellt.

**[0028]** Die Samples eines gegenwärtigen Fensters werden einer Zeit-Frequenz-Transformation unterzogen, die geeignete Konditionierungsoperationen, wie Vor-Filterung, Shading, usw., einschließen können (Schritt **206**). Irgendeine von verschiedenen Zeit-Frequenz-Transformationen, wie die Kurzzeit-Fourier-Transformation, die Filterbankanalyse, die diskrete Wavelet-Transformation, usw., kann verwendet werden.

**[0029]** Das Resultat der Zeit-Frequenz-Transformation ist, dass das anfängliche Zeitreihen-Eingangssignal  $x(t)$  in eine Zeit-Frequenzdarstellung  $X(f, i)$  transformiert wird, wobei  $t$  der Samplingindex zu der Zeitreihe  $x$  ist, und  $f$  und  $i$  sind jeweils die diskreten Variablen, welche die Frequenz- beziehungsweise Zeitdimensionen des Spektrogramms  $X$  indizieren. In der bevorzugten Ausführungsform wird der Logarithmus des Betrags von  $X$  anstelle von  $X$  in den folgenden Schritten, wenn es nicht anders angegeben ist, verwendet, d.h.

$$P(f, i) = 20 \log_{10} (|X(f, i)|).$$

**[0030]** Der Leistungspegel  $P(f, i)$  als Funktion der Zeit und der Frequenz wird von nun an als "Spektrogramm" bezeichnet.

**[0031]** Die Leistungspegel in den einzelnen Frequenzbändern  $f$  unterliegen sodann einer Hintergrundgeräuschschätzung (Schritt **208**). Ein Signaldetektor ermittelt das Vorhandensein eines Signals, das in dem stationären Hintergrundgeräusch verborgen ist (Schritt **210**) und überträgt nur die Spektrogramme, die ein Signal enthalten. Die Hintergrundgeräuschschätzung aktualisiert die Schätzung der Hintergrundgeräuschparameter, wenn kein Signal vorliegt.

**[0032]** Eine bevorzugte Ausführungsform für das Durchführen der Hintergrundgeräuschschätzung umfasst einen Leistungsdetektor, der die akustische Leistung in einem gleitenden Fenster für jedes Frequenzband  $f$  mittelt. Wenn die Leistung innerhalb einer vorbestimmten Zahl von Frequenzbändern eine Schwelle übersteigt, die als eine bestimmte Anzahl von Standardabweichungen über dem Hintergrundgeräusch definiert ist, erklärt der Leistungsdetektor das Vorhandensein eines Signals, wenn:

$$P(f, i) > B(f) + \sigma(f),$$

wobei  $B(f)$  die mittlere Hintergrundgeräuschleistung in dem Band  $f$  ist,  $\sigma$  die Standardabweichung des Ge-

räusche in demselben Band und  $c$  eine Konstante ist. In einer alternativen Ausführungsform braucht die Geräuschschätzung nicht dynamisch zu sein, sondern könnte einmalig gemessen werden (zum Beispiel, während des Bootens eines Computers, auf dem die Software der Erfindung läuft).

**[0033]** Die Spektrogramme, die durch den Signaldetektor geführt werden, werden dann einer harmonischen Detektorfunktion zugeführt (Schritt 212). Dieser Schritt ermöglicht dem System von Signalen zu unterscheiden, die nicht von der gleichen harmonischen Klasse wie das Eingangssignal sind, und für welche daher kein weiterer Vergleich notwendig ist. Zum Beispiel wird die menschliche Stimme durch das Vorhandensein eines Satzes Harmonischer zwischen 0,1 und ungefähr 3 kHz, mit einer Grundfrequenz (Pitch) von zwischen 90 Hz für erwachsene Männer bis zu 300 Hz für Kinder, gekennzeichnet.

**[0034]** Die Spektrogramme  $P$  von Schritt 206 werden sodann vorzugsweise umskaliert, so dass sie mit gespeicherten Vorlagen verglichen werden können (Schritt 214). Ein Verfahren zum Durchführen dieses Schrittes, besteht darin, jedes Element des Spektrogramms  $P(f, i)$  um eine Konstante  $k(i, m)$  nach oben zu verschieben, so dass die Differenz des quadratischen Mittelwerts von  $P(f, i) + k(i, m)$  und der  $m$ -ten  $T(f, m)$  minimiert wird. Dieses wird erreicht, indem man das folgende durchführt, wobei  $N$  die Zahl Frequenzbänder ist:

$$k(i, m) = \frac{1}{N} \sum_{j=1}^N [P(f, i) - T(f, m)].$$

**[0035]** In einer anderen Ausführungsform wird eine Gewichtung verwendet, um die Vorlagen vor dem Vergleich umzuskalieren. Die Gewichte  $w(i)$  sind zu dem SNR  $r(f, i)$  in dem Band  $f$  zu der Zeit  $i$  proportional, das als Differenz der Pegel, d.h.  $r(f, i) = P(f, i) - B(f)$ , für jedes Frequenzband berechnet wird. In dieser Ausführungsform wird jedes Element des Umskalierungsfaktors durch ein Gewicht gewichtet, das wie folgt definiert wird, wobei  $w_{\min}$  und  $w_{\max}$  voreingestellte Schwellen sind:

$$w(f, i) = w_{\min}, \text{ wenn } r(f, i) < w_{\min};$$

$$w(f, i) = w_{\max}, \text{ wenn } r(f, i) > w_{\max};$$

$$w(f, i) = r(f, i), \text{ sonst.}$$

**[0036]** In der bevorzugten Ausführungsform werden die Gewichte durch die Summe der Gewichte zu jedem Zeitframe normalisiert, d.h.:

$$w'(f, i) = w(f, i) / \text{Summe}_f(w(f, i)),$$

$$w'_{\min} = w_{\min} / \text{Summe}_f(w(f, i)),$$

$$w'_{\max} = w_{\max} / \text{Summe}_f(w(f, i)).$$

**[0037]** In diesem Fall ist die Umskalierungskonstante gegeben durch:

$$k(i, m) = \frac{1}{N} \sum_{j=1}^N [P(f, i) - T(f, m)] w'(f, i).$$

**[0038]** Der Effekt eines solchen Umskalierens besteht darin, vorzugsweise die Frequenzbänder der Vorlagen auszurichten, die ein höheres SNR haben. Jedoch ist das Umskalieren optional und braucht nicht in allen Ausführungsformen verwendet zu werden.

**[0039]** In einer anderen Ausführungsform wird das SNR der Vorlagen sowie das SNR der gemessenen Spektren für das Umskalieren der Vorlagen verwendet. Das SNR der Vorlage  $T(f, m)$  ist definiert durch  $r_N(f, m) = T(f, M) - B_N(f)$ , wobei  $B_N(f)$  das Hintergrundgeräusch in dem Frequenzband  $f$  zu der Zeit des Trainings ist. In einer Ausführungsform eines Gewichtungsschemas, das sowohl  $r$  als auch  $r_N$  verwendet, werden die Gewichte  $w_N$  als die Quadrat-Wurzel des Produktes der Gewichte für die Vorlagen und das Spektrogramm definiert:

$$w(f, i, m) = w_{\min}, \text{ wenn } \sqrt{r_N(f, m)r(f, i)} < w_{\min};$$

$$w(f, i, m) = w_{\max}, \text{ wenn } \sqrt{r_N(f, m)r(f, i)} > w_{\max};$$

$$w(f, i, m) = \sqrt{r_N(f, m)r(f, i)}, \text{ sonst.}$$

**[0040]** Andere Kombinationen von  $r_N$  und von  $r$  sind zulässig. In der bevorzugten Ausführungsform werden die Gewichte durch die Summe der Gewichte an jedem Zeitrahmen normalisiert, d.h.:

$$w'_2(f, i) = w_2(f, i)/\text{Summe}_f(w_2(f, i)),$$

$$w'_{\min} = w_{\min}/\text{Summe}_f(w_2(f, i)),$$

$$w'_{\max} = w_{\max}/\text{Summe}_f(w_2(f, i)),$$

**[0041]** Nach dem spektralen Umskalieren führt die bevorzugte Ausführungsform den Musterabgleich durch, um eine Vorlage  $T^*$  in einem Signalmodell zu finden, die am besten zu dem gegenwärtigen Spektrogramm  $P(f, i)$  passt (Schritt 216). Es besteht eine gewisse Breite in der Definition des Ausdrucks "am besten passen", sowie in dem Verfahren, das verwendet wird, um diese am besten Passende zu finden. In einer Ausführungsform wird die Vorlage mit der kleinsten Differenz des r.m.s. (quadratischen Mittels)  $d^*$  zwischen  $P + k$  und  $T^*$  aufgefunden. In der bevorzugten Ausführungsform wird der gemittelte r.m.s.-Abstand verwendet, wobei:

$$d(i, m) = \frac{1}{N} \sum_{f=1}^N [P(f, i) + k(i, m) - T(f, m)]^2 w'_2(f, i, m).$$

**[0042]** In dieser Ausführungsform tragen die Frequenzbänder mit dem geringsten SNR weniger zu der Abstands berechnung als jene Bänder mit einem größeren SNR bei. Die am besten passende Vorlage  $T^*(i)$  zu der Zeit  $i$  wird ausgewählt, indem man  $m$  so auffindet, dass  $d^*(i) = \text{Min}_m(d(i, m))$ .

**[0043]** Der letzte Bestandteil ist eine Klassifikationseinrichtung. Eine Bewertung wird für jede Klasse ange sammelt, und wenn genügend Daten gesammelt worden sind, wird eine Entscheidung getroffen. Zum Beispiel kann eine Bewertung der Mittelwert der Abstände  $d(i, m)$  über die Zeit  $i$  sein. In einer typischen Ausführungs form werden 8–20 Bewertungen angesammelt, von denen jede einem Puffer von stimmhafter Rede entspricht (im Gegensatz zu stimmloser Rede – Konsonanten –, da die Puffer ohne stimmhafte Rede nicht so viel Infor mationen hinsichtlich der Identität des Sprechers enthalten). Die Klassifikationsentscheidung kann schlicht das Vergleichen einer Bewertung mit einer Schwelle umfassen, was in einer binären Ermittlung resultiert, oder man kann eine weiche „Klassifikationseinrichtung“, wie ein neuronales Netz, verwenden. Alternativ kann eine Ent scheidung zu jeder möglichen gewünschten Zeit oder zu jedem gewünschten Ereignis erzwungen werden, und es wird die am besten passende Klasse zusammen mit der Bewertung an diesem Punkt wiedergegeben. Die Bewertung kann einen Bestandteil enthalten, der den Beitrag der Grundfrequenz mit der Gesamtbewertung in Bezug setzt. Die bevorzugte Ausführungsform dieses Bestandteils ist von der Form  $K(f_0 - f_{\text{Quelle}})^2$ , wobei  $f_0$  die gemessene Grundfrequenz ist,  $f_{\text{Quelle}}$  die Grundfrequenz des Quellenmodells ist, und  $K$  eine Proportionalitäts konstante ist.

**[0044]** Insbesondere ist in der bevorzugten Ausführungsform die Bewertung der Durchschnitt des Abstands über die Zeit, plus einem Grundfrequenzterm, d.h.

$$s = \frac{1}{N} \sum_{i=1}^{I+N} d^*(i) + K(f_0 - f_{\text{Quelle}})^2,$$

wobei der Durchschnitt über  $N$  Punkte beginnend zu der Zeit  $i = 1$  genommen wird. In diesem Fall muss die Bewertung  $s$  minimiert werden. Wenn  $s$  nicht einen gewählten Schwellenwert  $T_{\text{unbekannt}}$  für sämtliche Modelle erreicht, dann wird erklärt, dass die Quelle von einer unbekannten Art ist. Andernfalls wird erklärt, dass die Quelle zu der Klasse mit der niedrigsten Bewertung gehört.

**[0045]** Einzelne oder mehrfache Signalmodelle, von denen jedes eine oder mehrere Vorlagen enthält, können in verschiedenen Anwendungen angewendet werden, um ein akustisches Eingangssignal zu klassifizieren. In dem Fall eines einzelnen Signalmodells ist die Klassifikation binär.

#### Hintergrundgeräuschschätzung und Signaldetektion

**[0046]** [Fig. 3](#) ist ein Flussdiagramm, das eine ausführlichere Beschreibung des Prozesses der Hintergrundgeräuschschätzung und der Signaldetektion, die bereits kurz als Schritte **208** beziehungsweise **210** in [Fig. 2](#) beschrieben wurden, bereitstellt. Die Hintergrundgeräuschschätzung aktualisiert die Schätzungen der Hintergrundgeräuschparameter, wenn kein Signal vorliegt. Ein Signaldetektor ist nützlich, um gegenüber einem ununterbrochene Hintergrundgeräusch zu unterscheiden. Es ist wichtig, sicherzustellen, dass die Klassifikation auf nur einem Signal basiert und nicht durch das Hintergrundgeräusch beeinflusst wird.

**[0047]** Der Prozess beginnt in einem Anfangsprozesszustand (Schritt **302**). Der Prozess benötigt eine genügende Anzahl (z. B. über 1 Sekunde) von Samples des Hintergrundgeräusches, bevor er den Mittelwert und die Standardabweichung des Geräusches verwenden kann, um ein Signal zu ermitteln. Dementsprechend stellt die Routine fest, ob eine genügende Anzahl von Samples des Hintergrundgeräusches erhalten worden ist (Schritt **304**). Wenn nicht, wird die vorliegende Vorlage benutzt, um die Geräuschschätzung zu aktualisieren (Schritt **306**), und der Prozess wird beendet (Schritt **310**). In einer Ausführungsform des Hintergrundgeräusch-Aktualisierungsprozesses werden die Spektrogrammelemente  $P(f, i)$  in einem Ringpuffer gehalten und verwendet, um den Mittelwert  $B$  und die Standardabweichung  $\sigma$  in jedem Frequenzband  $f$  zu aktualisieren. Die Hintergrundgeräuschschätzung gilt als fertiggestellt, wenn der Index  $i$  größer als eine voreingestellte Schwelle ist.

**[0048]** Wenn die Hintergrundsamples bereit sind (Schritt **304**), dann wird eine Bestimmung ausgeführt, ob der Signalpegel  $P(f, i)$  eines gegenwärtigen Eingangssignalsamples in einigen der Frequenzbänder signifikant über dem Hintergrund liegt (Schritt **308**). In einer bevorzugten Ausführungsform zeigt der Bestimmungsschritt, wenn die Leistung in einer vorbestimmten Anzahl von Frequenzbändern größer als eine Schwelle ist, die als eine bestimmte Anzahl von Standardabweichungen über dem Hintergrundgeräusch-Mittelwertpegel bestimmt ist, an, dass die Leistungsschwelle überschritten worden ist, d.h. wenn

$$P(f, i) > B(f) + c \sigma(f),$$

wobei  $c$  eine empirisch vorbestimmte Konstante ist (Schritt **312**). Der Prozess wird dann beendet (Schritt **310**). Wenn in Schritt **308** kein Signal von hinreichender Leistung ermittelt wird, dann wird die Hintergrundgeräuschstatistik in Schritt **306** aktualisiert, und der Prozess wird beendet (Schritt **310**).

#### Harmonischer Detektor

**[0049]** [Fig. 4](#) ist ein Flussdiagramm, das eine ausführlichere Beschreibung des Prozesses der harmonischen Detektion bereitstellt, der kurz als Schritt **212** im [Fig. 2](#) beschrieben wurde. Der harmonische Detektor detektiert das Vorhandensein von Spitzen in dem Spektrum eines Eingangssignalsamples, die eine harmonische Relation untereinander aufweisen. Dieser Schritt ist häufig nützlich, da ein großer Anteil von Quellen von Interesse Spektren besitzt, die dadurch gekennzeichnet sind, dass sie ein harmonisches Verhältnis zwischen ihren Frequenzbestandteilen aufweisen.

**[0050]** Der Prozess beginnt an einem Anfangsprozesszustand (Schritt **402**). Das transformierte Spektrum eines Eingangssignalsamples wird auf lokale Spitzen für Frequenzen bis zu einer maximalen Frequenz von  $f_{\max}$  abgetastet, um eine Spalte „aufzugreifen“ (Schritt **404**). Eine lokale Spalte wird an  $P(f)$  erklärt, wenn  $P(f-1) < P(f) < P(f+1)$ . Die Spitzen, die mehr als eine Schwelle  $\epsilon$  über den benachbarten Spektrumswerten stehen, d.h. jene  $f$ , für die gilt  $P(f-1) + \epsilon < P(f) < P(f+1) + \epsilon$ , werden extrahiert (Schritt **406**). Jede jener Spitzen stellt eine „Wahlstimme“ für jede der Grundfrequenzen  $f_0$  dar (Schritt **408**). Die Schätzung der bevorzugten Ausführungsform von  $V_0(f_0)$  ist Untergrenze( $f_{\max}/f_0$ ). Da niedrigere Werte von  $f_0$  weniger Harmonische für ein gegebenes  $f_{\max}$  als höhere  $f_0$  aufweisen, werden die Wahlstimmen durch die erwartete Zahl von Harmonischen in dem Frequenzbereich in Betracht von  $V_0(f_0)$  normalisiert (Schritt **410**). Wenn das Verhältnis  $V(f_0)/V_0(f_0)$  größer als eine Schwelle ist (Schritt **412**), wird erklärt, dass ein harmonisches Verhältnis vorliegt.

#### Musterabgleich

**[0051]** [Fig. 5](#) ist ein Flussdiagramm, das eine ausführlichere Beschreibung des Prozesses des Musterab-

gleichs bereitstellt, der kurz als Schritt **216** von [Fig. 2](#) beschrieben wurde. Der Prozess beginnt an einem Anfangsprozesszustand (Schritt **502**). Der Musterabgleichprozess findet eine Vorlage  $T^*$  in dem Signalmodell, die am besten zu einem gegenwärtigen Spektrogramm  $P(f, i)$  passt (Schritt **504**). Der Musterabgleichprozess ist auch für den Lernprozess des Signalmodells verantwortlich. Es besteht eine gewisse Breite in der Definition des Ausdrucks "am besten passen", sowie in dem Verfahren, das verwendet wird, um diese am besten Passende zu finden. In einer Ausführungsform wird die Vorlage mit der kleinsten Differenz des r.m.s. (quadratischen Mittels)  $d^*$  zwischen  $P + k$  und  $T^*$  gefunden. In der bevorzugten Ausführungsform wird der gemittelte r.m.s.-Abstand verwendet, wobei:

$$d(i, m) = \frac{1}{N} \sum_{f=1}^N [P(f, i) + k(i, m) - T(f, m)]^2 w_2'(f, i, m).$$

**[0052]** In dieser Ausführungsform tragen die Frequenzbänder mit dem geringsten SNR weniger zu der Abstandsberechnung bei als jene Bänder mit einem größeren SNR. Die am besten passende Vorlage  $T^*(f, i)$ , d.h. die Ausgabe von Schritt **504** zu der Zeit  $i$ , wird ausgewählt, indem man  $m$  so auffindet, dass  $d^*(i) = \text{Min}_m[d(i, m)]$ . Wenn das System sich nicht in dem Lernmodus befindet (Schritt **506**), dann stellt  $T^*(f, i)$  auch die Ausgabe des Prozesses dar, insofern es die nächste Vorlage darstellt (Schritt **508**). Der Prozess wird sodann beendet (Schritt **510**).

**[0053]** Wenn das System in dem Lernmodus ist (Schritt **506**), wird die Vorlage  $T^*(f, i)$ , die  $P(f, i)$  am ähnlichsten ist, verwendet, um das Signalmodell anzupassen. Die Art, in der  $T^*(f, i)$  in dem Modell eingebaut wird, hängt von dem Wert von  $d^*(i)$  ab (Schritt **512**). Wenn  $d^*(i) < d_{\max}$ , wobei  $d_{\max}$  eine vorbestimmte Schwelle ist, wird dann  $T^*(f, i)$  angepasst (Schritt **516**), und der Prozess wird beendet (Schritt **510**). Die bevorzugte Ausführungsform von Schritt **516** wird so implementiert, dass  $T^*(f, i)$  der Durchschnitt aller Spektren  $P(f, i)$  ist, die verwendet werden, um  $T^*(f, i)$  zusammenzusetzen. In der bevorzugten Ausführungsform wird die Zahl  $n_m$  der Spektren, die mit  $T(f, m)$  assoziiert sind, im Speicher gehalten, und wenn ein neues Spektrum  $P(f, i)$  verwendet wird, um  $T(f, m)$  anzupassen, ist die angepasste Vorlage:

$$T(f, m) = [n_m T(f, m) + P(f, i)]/(n_m + 1)$$

und die Zahl der Muster, die der Vorlage  $m$  entsprechen, wird ebenso angepasst:

$$n_m = n_m + 1.$$

**[0054]** Wieder mit Bezug auf Schritt **512** wird, wenn  $d^*(i) > d_{\max}$ , dann eine neue Vorlage,  $T^*(f, i) = P(f, i)$  mit einem Gewicht  $n_m = 1$  erzeugt (Schritt **514**), und der Prozess wird beendet (Schritt **510**).

### Computerimplementierung

**[0055]** Die Erfindung kann in Hardware oder Software oder in einer Kombination von beiden (z. B. programmierbare Logikarrays) implementiert werden. Wenn es nicht anders angegeben ist, stehen die Algorithmen, die als Teil der Erfindung eingeschlossen werden, nicht mit irgendeinem bestimmten Computer oder einer anderen Vorrichtung inhärent in Verbindung. Insbesondere können verschiedene universelle Maschinen mit den Programmen benutzt werden, die in Übereinstimmung mit der Lehre hierin geschrieben werden, oder es kann geeigneter sein, spezialisierte Vorrichtungen zu konstruieren, um die erforderlichen Verfahrensschritte durchzuführen. Vorzugsweise wird jedoch die Erfindung in einem oder mehreren Computerprogrammen implementiert, die auf programmierbaren Systemen ausführbar sind, von denen jedes zumindest einen Prozessor, zumindest ein Datenspeichersystem (einschließlich flüchtigen und nicht-flüchtigen Speicher und/oder Speicherungselemente), zumindest eine Eingangseinrichtung und zumindest eine Ausgangseinrichtung umfasst. Jeder solcher programmierbaren Systembestandteile bildet eine Einrichtung zum Ausführen einer Funktion. Der Programmcode wird von den Prozessoren ausgeführt, um die hierin beschriebenen Funktionen auszuführen.

**[0056]** Jedes solcher Programme kann in irgendeiner gewünschten Computersprache (einschließlich von Maschinen-, Assembler-, Hoch-, oder objektorientierten Sprachen) implementiert sein, um mit dem Computersystem zu kommunizieren. In jedem Fall kann die Sprache eine Compiler- oder Interpreter-Sprache sein.

**[0057]** Jedes solcher Computerprogramme wird bevorzugt auf einem Speichermedium oder einer Speichereinrichtung (z. B. ROM, CD-ROM, oder ein magnetisches oder optisches Medium) gespeichert, das von einem allgemeinen oder spezialisierten programmierbaren Computer lesbar ist, um den Computer zu konfigurieren und zu betreiben, wenn das Speichermedium oder die Speichereinrichtung von dem Computer gelesen wird,

um die hierin beschriebenen Prozeduren auszuführen. Das erfindungsgemäße System kann ebenso als ein computerlesbares Speichermedium implementiert angesehen werden, das mit einem Computerprogramm konfiguriert ist, wobei das so konfigurierte Speichermedium einen Computer veranlasst, in einer speziellen und vorbestimmten Art zu arbeiten, um die hierin beschriebenen Funktionen auszuführen.

**[0058]** Es ist eine Zahl von Ausführungsformen der Erfindung beschrieben worden. Dennoch wird es verstanden werden, dass verschiedene Änderungen unternommen werden können, ohne von dem Bereich der Erfindung abzuweichen. Zum Beispiel können einige der Schritte von verschiedenen von den Algorithmen von der Reihenfolge unabhängig sein, und können somit in einer anderen Reihenfolge als der oben beschriebenen ausgeführt werden. Dementsprechend liegen andere Ausführungsformen innerhalb des Bereichs der folgenden Ansprüche.

### Patentansprüche

1. Ein Verfahren zur Klassifikation eines akustischen Signals innerhalb eines digitalisierten akustischen Eingangssignals, enthaltend:

- (a) Umwandeln des digitalisierten akustischen Eingangssignals in eine Zeit-Frequenz-Darstellung;
- (b) Detektieren des Vorhandenseins von Harmonischen in der Zeit-Frequenz-Darstellung;
- (c) Umskalieren der Zeit-Frequenz-Darstellung;
- (d) Vergleichen der umskalierten Zeit-Frequenz-Darstellung mit einer Mehrzahl von Zeit-Frequenz-Spektrogramm-Vorlagen von Signalmodellen, wenn Harmonische in der Zeit-Frequenz-Darstellung detektiert werden;
- (e) Bestimmen eines Abstands zwischen der umskalierten Zeit-Frequenz-Darstellung und jeder einer Mehrzahl von Zeit-Frequenz-Spektrogramm-Vorlagen von den Signalmodellen;
- (f) Bestimmen einer passenden Spektrogramm-Vorlage, die gemäß den Abständen am besten zu der umskalierten Zeit-Frequenz-Darstellung passt;
- (g) Zuweisen des digitalisierten akustischen Eingangssignals zu einem ersten Signalmodell, das die passende Spektrogramm-Vorlage enthält.

2. Das Verfahren von Anspruch 1, des weiteren den Schritt des Zurückweisens eines zweiten Signalmodells einschließend, das von der umskalierten Zeit-Frequenz-Darstellung zu weit entfernt ist.

3. Das Verfahren eines der obigen Ansprüche, des weiteren Aktualisieren der Hintergrundgeräuschstatistik umfassend, wenn der Leistungspegel eines Frequenzbands in der Zeit-Frequenz-Darstellung niedriger als eine Leistungsschwelle ist; und

worin ein Detektieren des Vorhandenseins von Harmonischen geschieht, wenn der Leistungspegel höher als eine Leistungsschwelle ist.

4. Das Verfahren eines der obigen Ansprüche, des weiteren umfassend:

Trainieren des ersten Signalmodells durch Erzeugen einer neuen Zeit-Frequenz-Spektrogramm-Vorlage für das erste Signalmodell auf der Grundlage der Zeit-Frequenz-Darstellung des Eingangssignals, wenn die Zeit-Frequenz-Darstellung des Eingangssignals um mehr als eine Schwelle von einer Vorlage von dem ersten Signalmodell, die der Zeit-Frequenz-Darstellung des Eingangssignals am ähnlichsten ist, verschieden ist; und/oder

Trainieren des ersten Signalmodells durch Mitteln der Zeit-Frequenz-Darstellung des Eingangssignals in eine vorhandene Zeit-Frequenz-Spektrogramm-Vorlage für das erste Signalmodell, wenn die Zeit-Frequenz-Darstellung des Eingangssignals sich um weniger als eine Schwelle von der vorhandenen Zeit-Frequenz-Spektrogramm-Vorlage unterscheidet.

5. Ein System zur Klassifikation eines akustischen Signals innerhalb eines digitalisierten akustischen Eingangssignals, enthaltend:

- a. eine Computereinrichtung zum Umwandeln des digitalisierten akustischen Eingangssignals in eine Zeit-Frequenz-Darstellung;
- b. eine Computereinrichtung zum Detektieren des Vorhandenseins von Harmonischen in der Zeit-Frequenz-Darstellung;
- c. eine Computereinrichtung zum Umskalieren der Zeit-Frequenz-Darstellung;
- d. eine Computereinrichtung zum Vergleichen der umskalierten Zeit-Frequenz-Darstellung mit einer Mehrzahl von Zeit-Frequenz-Spektrogramm-Vorlagen von Signalmodellen, wenn Harmonische in der Zeit-Frequenz-Darstellung detektiert werden;
- e. eine Computereinrichtung zum Bestimmen eines Abstands zwischen der umskalierten Zeit-Frequenz-Darstellung und jeder einer Mehrzahl von Zeit-Frequenz-Spektrogramm-Vorlagen von den Signalmodellen;

- f. eine Computereinrichtung zum Bestimmen einer passenden Spektrogramm-Vorlage, die gemäß den Abständen am besten zu der umskalierten Zeit-Frequenz-Darstellung passt;
- g. eine Computereinrichtung zum Zuweisen des digitalisierten akustischen Eingangssignals zu einem ersten Signalmodell, das die passende Spektrogramm-Vorlage enthält.

6. Das System von Anspruch 5, des weiteren eine Computereinrichtung zum Zurückweisen eines zweiten Signalmodells einschließend, das von der umskalierten Zeit-Frequenz-Darstellung zu weit entfernt ist.

7. Das System von Anspruch 5 oder 6, des weiteren umfassend:  
eine Computereinrichtung zum Aktualisieren der Hintergrundgeräuschstatistik, wenn der Leistungspegel eines Frequenzbands in der Zeit-Frequenz-Darstellung niedriger als eine Leistungsschwelle ist; und  
worin ein Detektieren des Vorhandenseins von Harmonischen geschieht, wenn der Leistungspegel höher als eine Leistungsschwelle ist.

8. Das System von einem der Ansprüche 5–7, des weiteren umfassend:  
eine Computereinrichtung zum Trainieren des ersten Signalmodells durch Erzeugen einer neuen Zeit-Frequenz-Spektrogramm-Vorlage für das erste Signalmodell auf der Grundlage der Zeit-Frequenz-Darstellung des Eingangssignals, wenn die Zeit-Frequenz-Darstellung des Eingangssignals um mehr als eine Schwelle von einer Vorlage von dem ersten Signalmodell, die der Zeit-Frequenz-Darstellung des Eingangssignals am ähnlichsten ist, verschieden ist; und/oder  
eine Computereinrichtung zum Trainieren des ersten Signalmodells durch Mitteln der Zeit-Frequenz-Darstellung des Eingangssignals in eine vorhandene Zeit-Frequenz-Spektrogramm-Vorlage für das erste Signalmodell, wenn die Zeit-Frequenz-Darstellung des Eingangssignals sich um weniger als eine Schwelle von der vorhandenen Zeit-Frequenz-Spektrogramm-Vorlage unterscheidet.

9. Ein Computerprogramm, das sich auf einem computerlesbaren Medium befindet, zur Klassifikation eines akustischen Signals innerhalb eines digitalisierten akustischen Eingangssignals, wobei das Computerprogramm Anweisungen umfasst, um den Computer zu veranlassen:

- (a) das digitalisierte akustische Eingangssignal in eine Zeit-Frequenz-Darstellung umzuwandeln;
- (b) das Vorhandensein von Harmonischen in der Zeit-Frequenz-Darstellung zu detektieren;
- (c) die Zeit-Frequenz-Darstellung umzuskalieren;
- (d) die umskalierte Zeit-Frequenz-Darstellung mit einer Mehrzahl von Zeit-Frequenz-Spektrogramm-Vorlagen von Signalmodellen zu vergleichen, wenn Harmonische in der Zeit-Frequenz-Darstellung detektiert werden;
- (e) einen Abstand zwischen der umskalierten Zeit-Frequenz-Darstellung und jeder einer Mehrzahl von Zeit-Frequenz-Spektrogramm-Vorlagen von den Signalmodellen zu bestimmen;
- (f) eine passende Spektrogramm-Vorlage, die gemäß den Abständen am besten zu der umskalierten Zeit-Frequenz-Darstellung passt, zu bestimmen;
- (g) das digitalisierte akustische Eingangssignal einem ersten Signalmodell, das die passende Spektrogramm-Vorlage enthält, zuzuweisen.

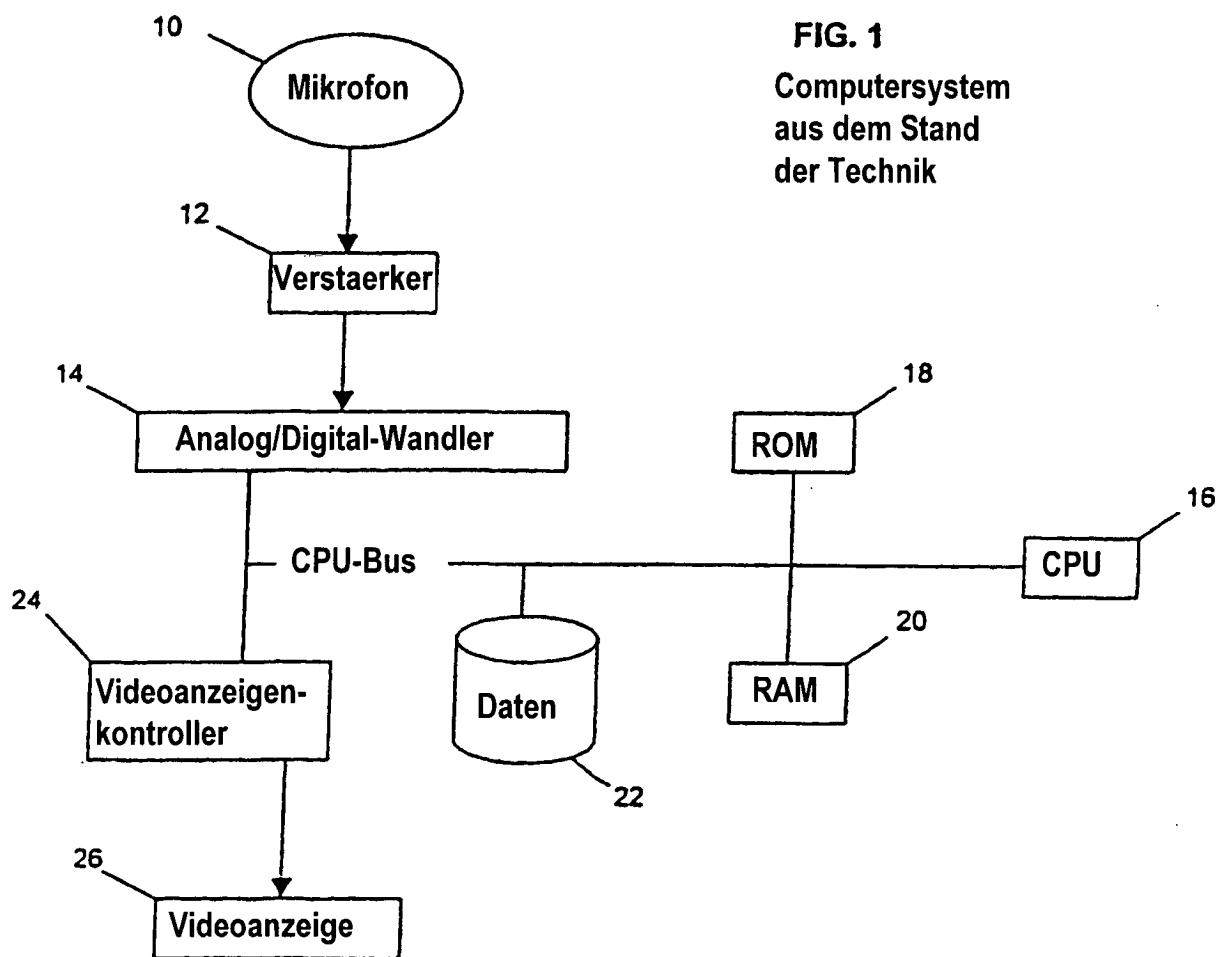
10. Das Computerprogramm von Anspruch 9, des weiteren Anweisungen zum Zurückweisens eines zweiten Signalmodells einschließend, das von der umskalierten Zeit-Frequenz-Darstellung zu weit entfernt ist.

11. Das Computerprogramm von Anspruch 9 oder 10, des weiteren Anweisungen zum Aktualisieren der Hintergrundgeräuschstatistik umfassend, wenn der Leistungspegel eines Frequenzbands in der Zeit-Frequenz-Darstellung niedriger als eine Leistungsschwelle ist; und  
worin ein Detektieren des Vorhandenseins von Harmonischen geschieht, wenn der Leistungspegel höher als eine Leistungsschwelle ist.

12. Das Computerprogramm von einem der Ansprüche 9–11, des weiteren Anweisungen umfassend zum Trainieren des ersten Signalmodells durch Erzeugen einer neuen Zeit-Frequenz-Spektrogramm-Vorlage für das erste Signalmodell auf der Grundlage der Zeit-Frequenz-Darstellung des Eingangssignals, wenn die Zeit-Frequenz-Darstellung des Eingangssignals um mehr als eine Schwelle von einer Vorlage von dem ersten Signalmodell, die der Zeit-Frequenz-Darstellung des Eingangssignals am ähnlichsten ist, verschieden ist; und/oder

zum Trainieren des ersten Signalmodells durch Mitteln der Zeit-Frequenz-Darstellung des Eingangssignals in eine vorhandene Zeit-Frequenz-Spektrogramm-Vorlage für das erste Signalmodell, wenn die Zeit-Frequenz-Darstellung des Eingangssignals sich um weniger als eine Schwelle von der vorhandenen Zeit-Frequenz-Spektrogramm-Vorlage unterscheidet.

Es folgen 5 Blatt Zeichnungen



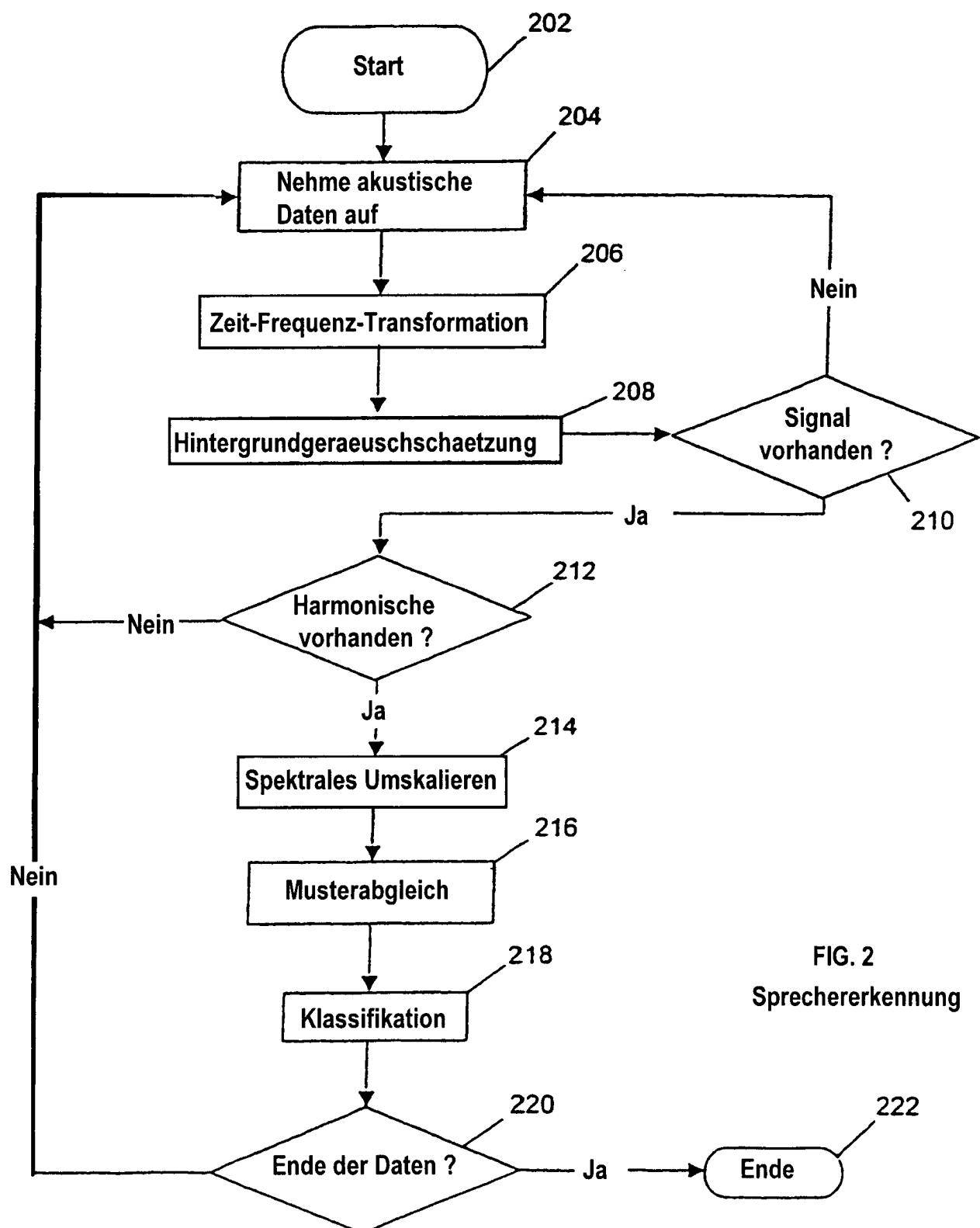


FIG. 2  
Sprechererkennung

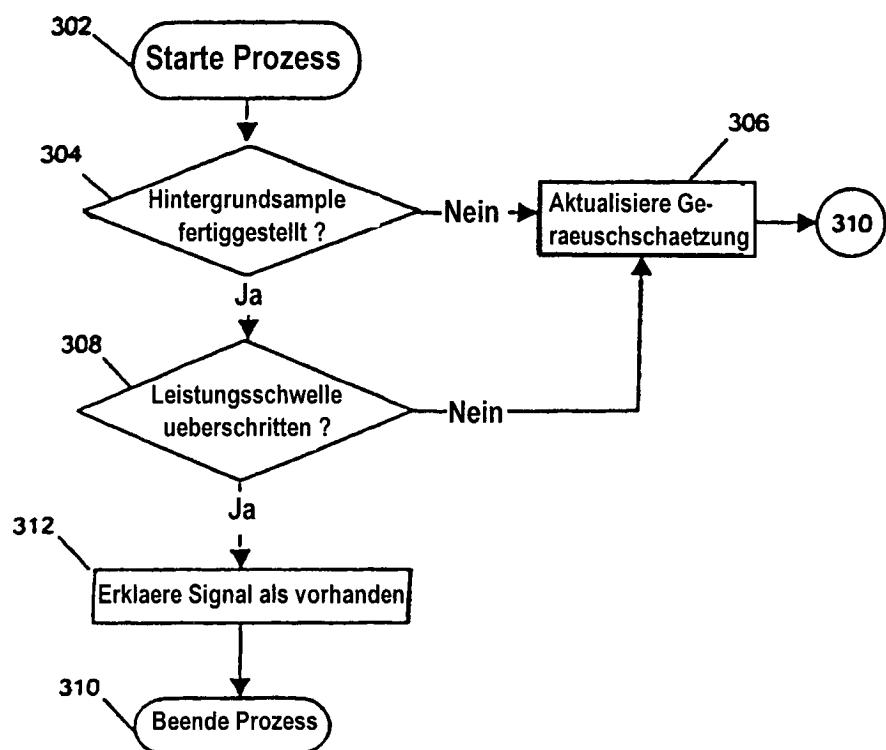
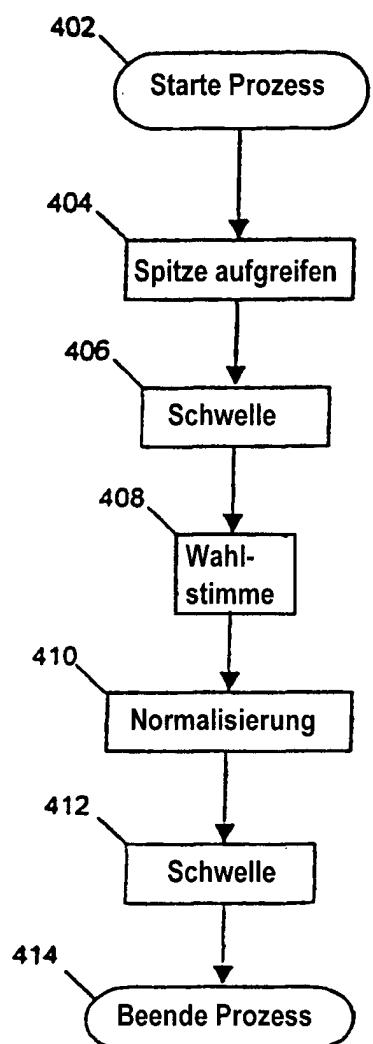


FIG. 3  
Hintergrundgeraeuschschaetzung  
und Signaldetektion



**FIG. 4**  
Harmonischer Detektor

**FIG. 5**  
Musterabgleich-Routine

