



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 697 25 172 T2 2004.04.08**

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 950 239 B1**
(21) Deutsches Aktenzeichen: **697 25 172.1**
(86) PCT-Aktenzeichen: **PCT/US97/04350**
(96) Europäisches Aktenzeichen: **97 915 156.0**
(87) PCT-Veröffentlichungs-Nr.: **WO 97/033273**
(86) PCT-Anmeldetag: **06.03.1997**
(87) Veröffentlichungstag
der PCT-Anmeldung: **12.09.1997**
(97) Erstveröffentlichung durch das EPA: **20.10.1999**
(97) Veröffentlichungstag
der Patenterteilung beim EPA: **24.09.2003**
(47) Veröffentlichungstag im Patentblatt: **08.04.2004**

(51) Int Cl.7: **G10L 15/20**
G10L 15/02

(30) Unionspriorität:
613269 08.03.1996 US

(73) Patentinhaber:
Motorola, Inc., Schaumburg, Ill., US

(74) Vertreter:
SCHUMACHER & WILLSAU,
Patentanwaltssozietät, 80335 München

(84) Benannte Vertragsstaaten:
AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI,
LU, MC, NL, PT, SE

(72) Erfinder:
FINEBERG, B., Adam, Saratoga, US

(54) Bezeichnung: **VERFAHREN UND GERÄT ZUM ERKENNEN VON GERÄUSCHSIGNALPROBEN AUS EINEM GERÄUSCH**

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 eingelegt, 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

Gebiet der Erfindung

[0001] Die vorliegende Erfindung bezieht sich im Allgemeinen auf Geräuscherkennung und im Besonderen auf Geräuscherkennung in einer Umgebung mit hohem oder variablem Rauschen.

Hintergrund der Erfindung

[0002] Die fortschreitende Technologie bewegt sich stetig in Richtung auf eine Kommerzialisierung von Geräuscherkennung durch elektronische Vorrichtungen, wie z. B. Spracherkennung. Im Allgemeinen gibt es zwei Typen von Spracherkennung. Ein Typ führt bestimmte Operationen aus, wenn der Anwender kurze Befehle gibt. Ein anderer Typ akzeptiert diktierete Sprache und gibt die Sprache als Text ein.

[0003] Die meisten Spracherkennung müssen durch den Anwender trainiert werden, bevor sie durch den Anwender gesprochene Wörter oder Phrasen erkennen können. Diese werden „sprecherabhängige“ Spracherkennung genannt, was bedeutet, dass der Spracherkennung durch die Stimme des Anwenders trainiert werden muss, bevor der Erkennung Anwenderwörter und -befehle interpretieren kann. Das Trainieren eines Spracherkennung erfordert, dass ein Anwender bestimmte Wörter oder Phrasen in den Erkennung spricht, üblicherweise viele Male, so dass der Spracherkennung das Sprachmuster des Anwenders erkennt. Später, wenn der Anwender den Spracherkennung verwendet, vergleicht der Spracherkennung das eingegebene Stimmensignal mit verschiedenen gespeicherten Sprachschablonen, um eine Schablone zu finden, die dem eingegebenen Stimmensignal am meisten gleicht. Dieses Verfahren wird „Musteranpassung“ genannt.

[0004] Ein Anwender „trainiert“ einen Spracherkennung im Allgemeinen in einer Umgebung, die ein relativ niedriges störendes Rauschen hat. Anschließend müssen die meisten Spracherkennung in Umgebungen mit niedrigem störenden Rauschen verwendet werden. Anderenfalls ist der Spracherkennung nicht in der Lage, gesprochene Wörter vom Hintergrundrauschen zu trennen. Da, wo Spracherkennung in einer Umgebung mit niedrigem Rauschen verwendet werden, wird eine ziemlich hohe Erkennungsrate erreicht. Wenn der Spracherkennung an einem Standort trainiert wird, der ein moderates, konstantes Hintergrundrauschen hat, und anschließend in einer Umgebung verwendet wird, die das selbe moderate, konstante Hintergrundrauschen hat, wird eine hohe Erkennungsrate erreicht. Wenn diese Spracherkennung jedoch in Umgebungen mit hohem Rauschen mit negativem Signal-Rausch-Verhältnis und Umgebungen, wo das vorhandene Rauschen von dem Hintergrundrauschen, das in der Trainingssitzung vorhanden ist, verschieden ist, verwendet wird, fällt die Erkennungsrate auf sehr niedrige, unbrauchbare Ge-

nautigkeitspegel.

[0005] Um das Problem des Hintergrundrauschens zu korrigieren, versuchen konventionelle Spracherkennung die Merkmale des umgebenden Rauschens einzuschätzen und bestimmen dann die Auswirkungen auf die Stimme des Anwenders. Verschiedene Techniken sind eingearbeitet, um statistische oder parametrische Modelle des Rauschens zu erzeugen, die von dem Geräuschsignal subtrahiert werden. In Umgebungen mit hohem und variablem Rauschen sind diese Modelle sehr ungenau.

[0006] Die DE-A-43 17 372 offenbart ein Verfahren zum Berechnen von Merkmalswerten eines akustischen Signals über eine Vielzahl von Frequenzbändern, einschließlich dem Bestimmen von minimalen und maximalen Merkmalswerten, um normalisierte Merkmalswerte zu bestimmen und zum Vergleichen des normalisierten Merkmalswertes mit Merkmalsvektorschablonen. Sie offenbart nicht das Ableiten von gewichteten minimalen und maximalen Merkmalswerten durch Verwenden einer Gewichtungsfunktion.

Zusammenfassung der Erfindung

[0007] In einem ersten Aspekt stellt die vorliegende Erfindung ein Verfahren gemäß Anspruch 1 zur Verfügung. In einem weiteren Aspekt stellt die vorliegende Erfindung einen Geräuscherkennung gemäß Anspruch 9 zur Verfügung.

Kurze Beschreibung der Zeichnungen

[0008] **Fig. 1** ist ein Blockdiagramm eines Stimmerkennungers gemäß einer bevorzugten Ausführungsform der vorliegenden Erfindung.

[0009] **Fig. 2** zeigt ein Flussdiagramm einer bevorzugten Ausführungsform der vorliegenden Erfindung.

[0010] **Fig. 3** zeigt ein Flussdiagramm eines Verfahrens, das verwendet wird, um Merkmalswerte gemäß einer bevorzugten Ausführungsform der vorliegenden Erfindung zu berechnen.

[0011] **Fig. 4** ist eine Darstellung eines Leistungsspektrums einer Geräuschsignalprobe mit darauf angelegten Frequenzfiltern gemäß der vorliegenden Erfindung.

[0012] **Fig. 5** zeigt eine Matrix von Merkmalen der Geräuschsignalprobe gemäß einer bevorzugten Ausführungsform der vorliegenden Erfindung.

[0013] **Fig. 6** zeigt eine Matrix von normalisierten Merkmalen für die Merkmale von **Fig. 3** gemäß einer bevorzugten Ausführungsform der vorliegenden Erfindung.

Beschreibung einer bevorzugten Ausführungsform

[0014] Eine bevorzugte Ausführungsform der vorliegenden Erfindung wird in einer widerstandsfähigen Stimmerkennung für Geräuscherkennung verwendet. Die bevorzugte Ausführungsform ist zur Verwendung

in Zellulartelefonen in Automobilen gut geeignet, wo ein Anwender beide Hände am Lenkrad belassen kann, mit dem Blick auf die Straße, und doch ein Telefongespräch führen kann, sogar bei heruntergelassenen Fenstern und lautgestelltem Stereosystem. Anders als konventionelle Spracherkennung, die eine unbrauchbare, schlechte Genauigkeitsrate unter hohen und/oder variablen Rauschbedingungen haben, sind Geräuscherkennung, die gemäß der bevorzugten Ausführungsform der vorliegenden Erfindung konstruiert sind, widerstandsfähig und können in Umgebungen, die variables Rauschen und Rauschpegel haben, die größer sind als die Lautstärke der Sprache des Anwenders, eine sehr hohe Genauigkeitsrate erreichen.

[0015] Die vorliegende Erfindung wird im Folgenden in Verbindung mit den Zeichnungen beschrieben. Im Besonderen wird die bevorzugte Ausführungsform mit Bezug auf **Fig. 1** in Kombination mit den anderen Abbildungen beschrieben.

[0016] Die vorliegende Erfindung kann auf die Erkennung eines beliebigen akustischen Geräusches angewendet werden. Das akustische Geräusch kann z. B. Sprache, Grunzgeräusche, Geräusche, die von Tieren gemacht werden, Geräusche, die durch Instrumente, einschließlich Schlaginstrumente, gemacht werden oder ein beliebiger anderer Typ von Geräusch sein. Am häufigsten bezieht sich die vorliegende Erfindung auf das Erkennen von Sprache.

[0017] **Fig. 1** zeigt einen Geräuscherkennung 100 gemäß einer bevorzugten Ausführungsform der vorliegenden Erfindung. In der bevorzugten Ausführungsform wird ein akustisches Signal in einem Analog-Digital-Wandler (ADC) 105 des Geräuscherkenners 100 eingegeben, wo das Signal in ein digitales Signal umgewandelt und bei einer Rate von 16 KHz abgetastet wird. Wenn angemessen, können andere Abtastraten verwendet werden, wie z. B. 8 KHz.

[0018] Die abgetasteten digitalen Signale werden in die Merkmalsvektorvorrichtung 110 eingegeben, die die abgetasteten digitalen Signale in Analyserahmen teilt. Jeder Analyserahmen kann so gewählt werden, dass er entweder eine feste Zeitbreite (wie z. B. 20 ms) oder eine variierte Zeitbreite hat, in Abhängigkeit von Signalmerkmalen wie z. B. Tonhöhe- bzw. Tonstärkeperioden ("pitch periods") oder anderen bestimmenden Faktoren. Der Startpunkt eines jeden Rahmens kann so gewählt werden, dass er entweder vor, an oder nach dem Endpunkt des vorherigen Rahmens liegt. In der bevorzugten Ausführungsform werden die Analyserahmen so gewählt, dass sie eine feste Zeitbreite haben, und jeder Analyserahmen beginnt an dem Endpunkt des vorherigen Analyserahmens.

[0019] Für jeden der Analyserahmen berechnet die Merkmalsvektorvorrichtung 110 einen Merkmalsvektor (**210** des Flussdiagramms von **Fig. 2**). Für jede beliebige Anzahl von Analyserahmen erzeugt die Merkmalsvektorvorrichtung 110 eine gleiche Anzahl von Merkmalsvektoren. Ein Merkmalsvektor ist eine

Reihe von Werten oder eine Vielzahl von Merkmalswerten, die von der Geräuschsignalprobe in einem gegebenen Analyserahmen abgeleitet werden. Diese Merkmalswerte stellen die in der Geräuschsignalprobe enthaltene Information dar.

[0020] Es gibt viele dem Fachmann bekannte Techniken zur Spracherkennung, die verwendet werden können, um Merkmalsvektoren zu bestimmen. Die Techniken umfassen lineare Prädiktionskodierungskoeffizienten (LPC-Koeffizienten), Cepstral-Koeffizienten, Log-Area-Verhältnisse, und Mel-Scale-Filterbank-Koeffizienten. Die bevorzugte Ausführungsform der vorliegenden Erfindung verwendet das Mel-Scale-Filterbankkoeffizienten-Verfahren, obwohl die vorliegende Erfindung mit anderen Merkmalsvektortechniken arbeitet, wie z. B. den oben aufgelisteten.

[0021] Mel-Scale-Filterbank-Koeffizienten werden folgendermaßen mit Bezug auf das Flussdiagramm von **Fig. 3** berechnet.

1. Die Geräuschsignalproben für einen Analyserahmen werden durch ein vorverzerrendes Hochfrequenzfilter geleitet, um das Spektrum der Geräuschsignalproben weiß zu machen (310 des Flussdiagramms von **Fig. 3**). Dies erhöht die relative Energie in den Hochfrequenzkomponenten im Vergleich zu der Energie der Niederfrequenzkomponenten. Man erhält Vorteile, wenn die bevorzugte Ausführungsform der vorliegenden Erfindung mit Sprachsignalen verwendet wird, weil Niederfrequenzkomponenten von Sprache eine relative Energie haben, die größer ist als die der Hochfrequenzkomponenten und weil die zwei Komponenten in dem vorverzerrenden Filter neu abgeglichen werden. In der bevorzugten Ausführungsform wird das Filtern gemäß der folgenden Gleichung erreicht:

$$p_i(k) = s_i(k) - s_i(k - 1)$$

wo $s_i(k)$ die Geräuschsignalprobe an der Position k in dem Analyserahmen „ i “ ist, $s_i(k - 1)$ die Geräuschsignalprobe in dem Analyserahmen „ i “ an der vorherigen Position in der Zeit „ $k - 1$ “ ist und $p_i(k)$ die vorverzerrte Geräuschsignalprobe an der Position „ k “ in dem Analyserahmen „ i “ ist. Dem Fachmann auf dem Gebiet der Spracherkennung ist klar, dass andere vorverzerrende Filter verwendet werden können.

2. Die vorverzerrten Geräuschsignalproben für jeden Analyserahmen werden durch eine Reihe von Filtern, die verschiedene Frequenzbänder abdecken, bandpassgefiltert. Die Filter können auf jede beliebige numerische Art angewendet werden, die entweder in der Zeitdomäne oder der Frequenzdomäne erforderlich ist. In der bevorzugten Ausführungsform werden die Filter in der Frequenzdomäne angewendet. Zuerst muss jedoch ein Leistungsspektrum der vorverzerrten Geräuschsignalprobe in dem Analyserahmen berechnet werden (320 von **Fig. 3**). Das Leistungsspektrum er-

hält man durch:

- a. Die vorverzerrten Geräuschsignalproben in dem Analyserahmen werden mit den Proben einer Fensterfunktion oder Gewichtungsfunktion multipliziert. Jede beliebige Fensterfunktion kann angewendet werden. Zum Zwecke des Erklärens der vorliegenden Erfindung gehen wir von einem einfachen rechteckigen Fenster aus (das Fenster hat einen Wert von 1.0 für alle Proben).
- b. Die Fourier-Transformation der vorverzerrten Geräuschsignalproben in jedem gefensterten Analyserahmen wird berechnet.
- c. Werte für das Leistungsspektrum werden durch Quadrieren der Fourier-Transformationswerte erhalten.

[0022] Nachdem die Werte für das Leistungsspektrum bestimmt sind, werden die Bandpassfilter in der Frequenzdomäne durch einen Filtergewichtungswert für jeden der Leistungsspektrumswerte angewendet (330 von Fig. 3). Obwohl viele Filtergewichtungsfunktionen in den Bandpassfiltern verwendet werden können, enthält die bevorzugte Ausführungsform ein Raised-Cosine-Gewichtungsprofil, das man in Fig. 4 sehen kann.

[0023] Fig. 4 zeigt ein Leistungsspektrum 400, das ein darauf aufgezwungenes Raised-Cosine-Profil 410 hat. Die Frequenzbänder für jedes Bandpassfilter, oder Raised-Cosine-Profil 410, werden in der bevorzugten Ausführungsform der vorliegenden Erfindung entlang der Frequenzachse gemäß einer Mel- oder Barkskala angeordnet, die der Frequenzantwort des menschlichen Ohres in etwa entspricht. Die Frequenzbänder für die Bandpassfilter (die Raised-Cosine-Profile 410) sind von 0 bis 1 KHz mit ungefähr linearen Abständen und oberhalb von 1 KHz mit logarithmischen Abständen angeordnet. Es können auch andere als die für die bevorzugte Ausführungsform definierten Filterabstände verwendet werden. Wie in Fig. 4 für die bevorzugte Ausführungsform zu sehen, überlappen sich die Bandpassfilter, oder Raised-Cosine-Profile 410. Die Ausgaben der Bandpassfilter oder Raised-Cosine-Profile 410, werden berechnet gemäß:

$$f_{ij} = \sum P_i(\omega) B_j(\omega)$$

wo $P_i(\omega)$ der Leistungsspektrumswert für den Analyserahmen „i“ bei der Frequenz ω ist, $B_j(\omega)$ die Bandpassfiltergewichtungsfunktion oder Frequenzantwort für den Filter „j“ bei der Frequenz ω ist, \sum die Additionsoperation über alle Frequenzen ω darstellt und f_{ij} die Bandpassfilterausgabe für den Analyserahmen „i“ und das Bandpassfilter „j“ ist.

[0024] Nachdem alle die Bandpassfilterausgaben für jeden Analyserahmen „i“ ($0 \leq i \leq n$) und jedes Bandpassfilter „j“ ($0 \leq j \leq m$) berechnet worden sind, berechnet die Merkmalsvektorvorrichtung 110 die Merkmalswerte „ v_{ij} “ der Geräuschsignalprobe durch Logarithmieren eines jeden Bandpassfilters f_{ij} (340

von Fig. 3). Das Ergebnis kann als eine Matrix gezeigt werden, wie z. B. die in Fig. 5 dargestellte Matrix, die mit „i“ Analyserahmen und „j“ Bandpassfiltern strukturiert ist, die eine Dimension von $n \times m$ hat. Alle die Merkmalswerte in einem Analyserahmen, v_{i1} bis v_{im} , bilden einen einzelnen Merkmalsvektor (wie in v_{i1} bis v_{ij} , Punkt 510), und alle Analyserahmen, $0 \leq i \leq n$, bilden eine Vielzahl von Merkmalsvektoren für die Geräuschsignalprobe.

[0025] Wenn die Vielzahl von Merkmalsvektoren für die Analyserahmen „i“ = 1 bis „n“ einmal berechnet worden sind, überprüft eine Min/Max-Vorrichtung 120 (Fig. 1), die mit der Merkmalsvektorvorrichtung 110 gekoppelt oder in ihr enthalten ist, alle Merkmalswerte in einem Frequenzband für das Bandpassfilter „j“ und findet den minimalen (\min_j) Merkmalswert und den maximalen (\max_j) Merkmalswert für das Frequenzband „j“ für alle Analyserahmen, $0 \leq i \leq n$ (220 von Fig. 2). Diese Minimal- und Maximalwerte werden verwendet, um die normalisierten Merkmalswerte „ v_{\geq} “ zu bestimmen.

[0026] Der Normalisierer 130 von Fig. 1 ist mit der Min/Max-Vorrichtung 120 und der Merkmalsvektorvorrichtung 110 gekoppelt. Der Normalisierer 130 normalisiert jeden der Merkmalswerte über ein Frequenzband oder Bandpassfilter, „j“, mit den gewichteten minimalen und maximalen Merkmalswerten für dieses Bandpassfilter, um die normalisierten Merkmalswerte „ v_{\geq} “ zu bestimmen (230 von Fig. 2).

[0027] Die Min/Max-Vorrichtung 120 findet einen gewichteten minimalen ($\omega_{\min j}$) Merkmalswert und einen gewichteten maximalen ($\omega_{\max j}$) Merkmalswert für das Frequenzband „j“ für alle Analyserahmen, $0 \leq i \leq n$ (220 von Fig. 2). Diese gewichteten Minimal- und Maximalwerte werden wie folgt berechnet:

$$\omega_{\min j} = \sum_{k=-r}^r \text{weight}[k] \min[j+k]$$

$$\omega_{\max j} = \sum_{k=-r}^r \text{weight}[k] \max[j+k]$$

wo r üblicherweise ein kleiner Wert wie z. B. 2 oder 3 ist, das Gewicht $[k]$ typischerweise eine Gewichtungsfunktion ist, wo der Mittelpunkt, Gewicht $[0]$, einen Wert hat, der größer oder gleich dem Wert aller anderen Gewichtungswerte ist, und

$$\sum_{k=-r}^r \text{weight}[k] = 1.0.$$

$$k = -r$$

[0028] Diese gewichteten Minimal- und gewichteten Maximalwerte werden verwendet, um die normalisierten Merkmalswerte „ v_{ij} “ zu bestimmen. In dieser Ausführungsform ist die Normalisierungsgleichung:

$$v_{ij} = (v_{ij} - \omega_{\min j}) / (\omega_{\max j} - \omega_{\min j})$$

wo v_{ij} einer der normalisierten Merkmalswerte ist, v_{ij} einer der Merkmalswerte ist, $\omega_{\min j}$ der gewichtete minimale Merkmalswert für das „j-te“ Frequenzband ist und $\omega_{\max j}$ der gewichtete maximale Merkmalswert für das „j-te“ Frequenzband ist.

[0029] Das Resultat des Normalisierungsprozesses kann als eine Matrix gezeigt werden, wie in **Fig. 6** dargestellt. Jeder der Analyserahmen „i“ von **Fig. 6** stellt einen normalisierten Merkmalsvektor (**610**) dar.

[0030] Der Komparator 140 von **Fig. 1** ist mit dem Normalisierer 130 gekoppelt und vergleicht die normalisierten Merkmalsvektoren mit Merkmalsvektorschablonen, um zu bestimmen, welche der Merkmalsvektorschablonen den normalisierten Merkmalsvektoren am meisten ähneln. Sätze von Merkmalsvektorschablonen, die Phrasen oder Befehle darstellen, werden in der Merkmalsvektorschablonenbibliothek 150 gespeichert. Der Komparator 140 vergleicht der Reihe nach die normalisierten Merkmalsvektoren von dem Normalisierer 130 mit jedem der Merkmalsvektorschablonen in der Merkmalsvektorschablonenbibliothek 150 (**240** von **Fig. 2**) und bestimmt, welcher Satz von den Merkmalsvektorschablonen [den normalisierten Merkmalsvektoren am meisten ähnelt. Dies wird] durch Berechnen eines Abstandsmaßes zwischen den normalisierten Merkmalsvektoren und jedem der Sätze von Merkmalsvektorschablonen [durchgeführt]. (Anmerkung des Übersetzers: Die eckig geklammerten Textbestandteile fehlen im "Druckexemplar".) Der Satz von Merkmalsvektorschablonen, der das minimale Abstandsmaß hat, wird als derjenige bestimmt, der den normalisierten Merkmalsvektoren am meisten ähnelt. Der Komparator 140 von **Fig. 1** gibt als eine am besten passende Anpassung den Satz von Merkmalsvektorschablonen aus der Merkmalsvektorschablonenbibliothek 150 aus, der den normalisierten Merkmalsvektoren (**250** von **Fig. 2**) am meisten ähnelt (der das minimale Abstandsmaß hat).

[0031] Es gibt mehrere gut bekannte Verfahren, durch die die Vielzahl von normalisierten Merkmalsvektoren mit den Merkmalsvektorschablonen verglichen werden können, um eine am besten passende Anpassung zu finden. Studien, die die bevorzugte Ausführungsform der vorliegenden Erfindung verwenden, zeigen, dass das Vergleichen der Vielzahl von normalisierten Merkmalsvektoren mit den Merkmalsvektorschablonen in einem dynamischen Zeitverwerfungsprozess die besten Resultate erzielt.

[0032] Wie zuvor erwähnt, ist die vorliegende Erfindung, wenn sie in einem sprecherabhängigen Geräuscherkennungssystem mit kleinem Vokabular verwendet wird, sehr widerstandsfähig und erhöht die

Erkennungsgenauigkeit in Umgebungen mit hohem und variablem Rauschen von unbrauchbaren Genauigkeitsraten auf sehr hohe Genauigkeitsraten.

[0033] Es sei darauf hingewiesen, dass die vorliegende Erfindung in vielen verschiedenen Geräuscherkennungssystemen verwendet werden kann. Alle solche unterschiedlichen Verwendungen werden durch die vorliegende Erfindung berücksichtigt.

Patentansprüche

1. Verfahren, das die folgenden Schritte umfasst: Berechnen von Merkmalswerten für eine Anzahl von Analyserahmen eines akustischen Signals, wobei die Merkmalswerte über eine Vielzahl von Frequenzbändern berechnet werden;
für jedes aus der Vielzahl von Frequenzbändern, Bestimmen in einer logarithmischen Domäne, welcher der Merkmalswerte in einem entsprechenden der Vielzahl von Frequenzbändern ein minimaler Merkmalswert über die Zeit ist und welcher der Merkmalswerte in dem entsprechenden der Vielzahl von Frequenzbändern ein maximaler Merkmalswert über die Zeit ist;
gekennzeichnet durch Gewichten der minimalen Merkmalswerte und der maximalen Merkmalswerte, um gewichtete minimale Merkmalswerte und gewichtete maximale Merkmalswerte zu erhalten;
in jedem Frequenzband, Vergleichen jedes der Merkmalswerte mit dem gewichteten minimalen Merkmalswert und dem gewichteten maximalen Merkmalswert des entsprechenden Frequenzbandes, um normalisierte Merkmalswerte zu erhalten, die in jedem der Frequenzbänder über die Zeit normalisiert sind, wobei alle die normalisierten Merkmalswerte für eine gegebene Anzahl von Analyserahmen einen aus einer Vielzahl von normalisierten Merkmalsvektoren definieren;
Vergleichen der Vielzahl von normalisierten Merkmalsvektoren mit Sätzen von Merkmalsvektorschablonen, um Merkmalsvektorschablonen zu bestimmen, die der Vielzahl von normalisierten Merkmalsvektoren am meisten ähneln; und
Ausgeben der Merkmalsvektorschablonen, die der Vielzahl von normalisierten Merkmalsvektoren am meisten ähneln.

2. Verfahren gemäß Anspruch 1, bei dem der Schritt des Berechnens der Merkmalswerte umfasst: Teilen eines Leistungsspektrums eines akustischen Geräuschsignals in die Vielzahl von Frequenzbändern;
Gewichten des Leistungsspektrums in jedem aus einer Vielzahl von Frequenzbändern gemäß einer Gewichtungsfunktion, um Filterausgaben zu erreichen; und
Berechnen der Merkmalswerte aus den Filterausgaben.

3. Verfahren gemäß Anspruch 2, bei dem das

Leistungsspektrum in einer Frequenzdomäne gemäß einem Raised-Cosine-Profil gewichtet wird.

4. Verfahren gemäß Anspruch 3, bei dem das Raised-Cosine-Profil die Filterausgaben gemäß einer Gleichung berechnet:

$$f_{ij} = SP_i(w)B_j(w)$$

wo f_{ij} eine der Filterausgaben an einem Analyserahmen „i“ aus einer Anzahl von Analyserahmen und einem Frequenzband „j“ aus einer Vielzahl von Frequenzbändern ist, $P_i(w)$ ein Leistungsspektrumswert für den Analyserahmen „i“ bei der Frequenz w ist, $B_j(w)$ eine Bandpassfiltergewichtungsfunktion, oder Frequenzantwort, für das Frequenzband „j“ bei der Frequenz w ist und S eine Additionsoperation über alle Frequenzen w darstellt.

5. Verfahren gemäß Anspruch 1, bei dem der Schritt des Vergleichens der normalisierten Merkmalsvektoren mit Sätzen von Merkmalsvektorschablonen, um die Merkmalsvektorschablonen zu bestimmen, die den normalisierten Merkmalsvektor am meisten ähneln, ein Vergleichen der normalisierten Merkmalsvektoren mit dem Satz von Merkmalsvektorschablonen in einem dynamischen Zeitverwertungsprozess umfasst.

6. Verfahren gemäß Anspruch 2, bei dem der Schritt des Berechnens der Merkmalswerte aus den Filterausgaben das Logarithmieren jeder Filterausgabe umfasst.

7. Verfahren gemäß Anspruch 2, bei dem das Leistungsspektrum in eine Vielzahl von Frequenzbändern gemäß einer Mel- oder Bark-Skala geteilt wird, um einer Frequenzantwort eines menschlichen Ohres in etwa zu entsprechen.

8. Verfahren gemäß Anspruch 1, bei dem jeder der normalisierten Merkmalsvektoren gemäß einer Gleichung gefunden wird:

$$v_{ij} \sim = (v_{ij} - \omega_{\min j}) / (\omega_{\max j} - \omega_{\min j})$$

wo v_{ij} eines der normalisierten Merkmale eines Analyserahmens „i“ aus der Anzahl von Analyserahmen und eines Frequenzbandes „j“ aus der Vielzahl von Frequenzbändern ist;

v_{ij} einer der Merkmalswerte an dem Analyserahmen „i“ und dem Frequenzband „j“ ist;

$\omega_{\min j}$ der gewichtete minimale Merkmalswert für das „j“-Frequenzband ist; und

$\omega_{\max j}$ der gewichtete maximale Merkmalswert für das „j“-Frequenzband ist.

9. Geräuscherkenner, der folgendes umfasst: eine Merkmalsvektorvorrichtung, die Merkmalswerte für eine Anzahl von Analyserahmen einer akusti-

schen Signaleingabe in den Geräuscherkenner berechnet, wobei die Merkmalswerte für eine Vielzahl von Frequenzbändern berechnet werden; eine Min/Max-Vorrichtung, die mit der Merkmalsvektorvorrichtung gekoppelt ist, um in einer logarithmischen Domäne zu bestimmen, welcher der Merkmalswerte in einem entsprechenden aus der Vielzahl von Frequenzbändern ein minimaler Merkmalswert über die Zeit und welcher der Merkmalswerte in dem entsprechenden aus der Vielzahl von Frequenzbändern ein maximaler Merkmalswert über die Zeit für jedes aus der Vielzahl von Frequenzbändern über die gesamte Anzahl von Analyserahmen ist; dadurch gekennzeichnet, dass die Min/Max-Vorrichtung die minimalen Merkmalswerte und die maximalen Merkmalswerte gewichtet, um gewichtete minimale Merkmalswerte und gewichtete maximale Merkmalswerte zu erhalten; einen mit der Min/Max-Vorrichtung gekoppelten Normalisierer, der in jedem aus der Vielzahl von Frequenzbändern jeden der Merkmalswerte mit dem gewichteten minimalen Merkmalswert und dem gewichteten maximalen Merkmalswert des entsprechenden aus der Vielzahl von Frequenzbändern vergleicht, um normalisierte Merkmalswerte, normalisiert über die Zeit in jedem der Frequenzbänder, zu erhalten, bei dem alle von den normalisierten Merkmalswerten für einen bestimmten der Analyserahmen einen aus einer Vielzahl von normalisierten Merkmalsvektoren bestimmen, und einen mit dem Normalisierer gekoppelten Komparator, der die Vielzahl von normalisierten Merkmalsvektoren mit Sätzen von Merkmalsvektorschablonen vergleicht, um Merkmalsvektorschablonen zu bestimmen, die der Vielzahl von normalisierten Merkmalsvektoren am meisten ähneln, wobei der Komparator die Merkmalsvektorschablonen ausgibt, die der Vielzahl von normalisierten Merkmalsvektoren am meisten ähneln.

10. Geräuscherkenner gemäß Anspruch 9, bei dem der Normalisierer jeden aus der Vielzahl der normalisierten Merkmalsvektoren gemäß der folgenden Gleichung berechnet:

$$v_{ij} \sim = (v_{ij} - \omega_{\min j}) / (\omega_{\max j} - \omega_{\min j})$$

wo v_{ij} einer der normalisierten Merkmalswerte eines Analyserahmens „i“ aus der Anzahl von Analyserahmen und eines Frequenzbandes „j“ aus der Vielzahl von Frequenzbändern ist;

v_{ij} einer von den Merkmalswerten an dem Analyserahmen „i“ und dem Frequenzband „j“ ist;

$\omega_{\min j}$ der gewichtete minimale Merkmalswert für das „j“-Frequenzband ist; und

$\omega_{\max j}$ der gewichtete maximale Merkmalswert für das „j“-Frequenzband ist.

11. Geräuscherkenner gemäß Anspruch 9, bei dem der Komparator mit einer Merkmalsvektorschab-

blonenbibliothek gekoppelt ist, wobei die Merkmalsvektorschablonenbibliothek die Sätze von Merkmalsvektorschablonen enthält.

Es folgen 3 Blatt Zeichnungen

Anhängende Zeichnungen

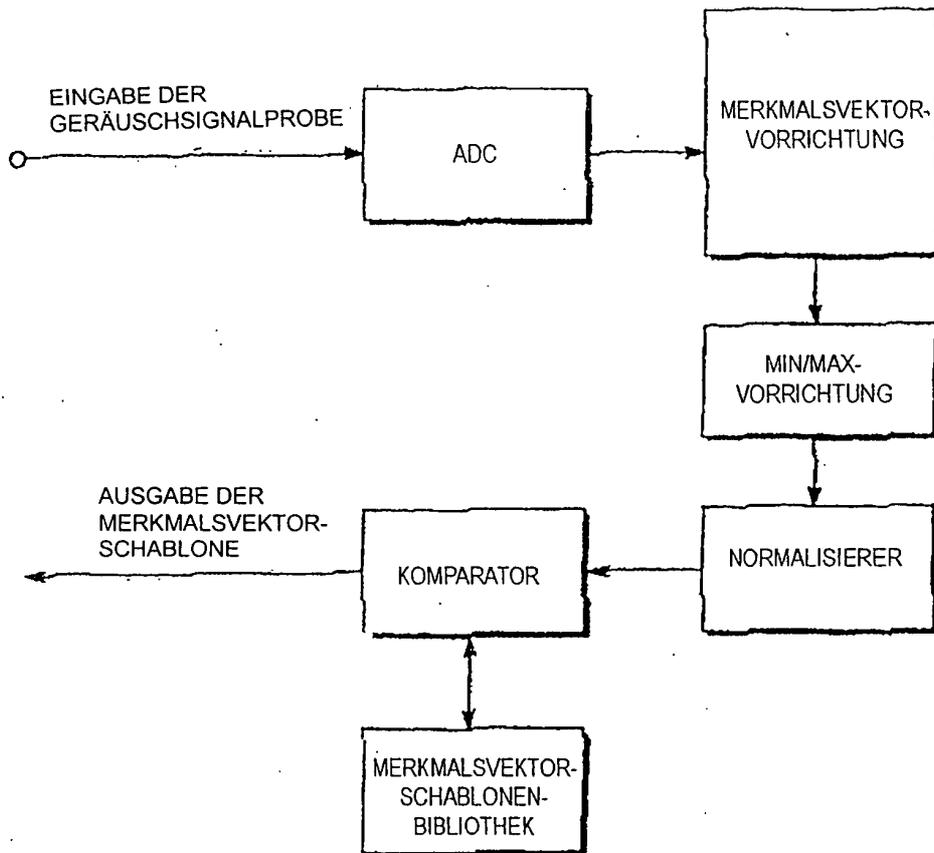


FIG.1

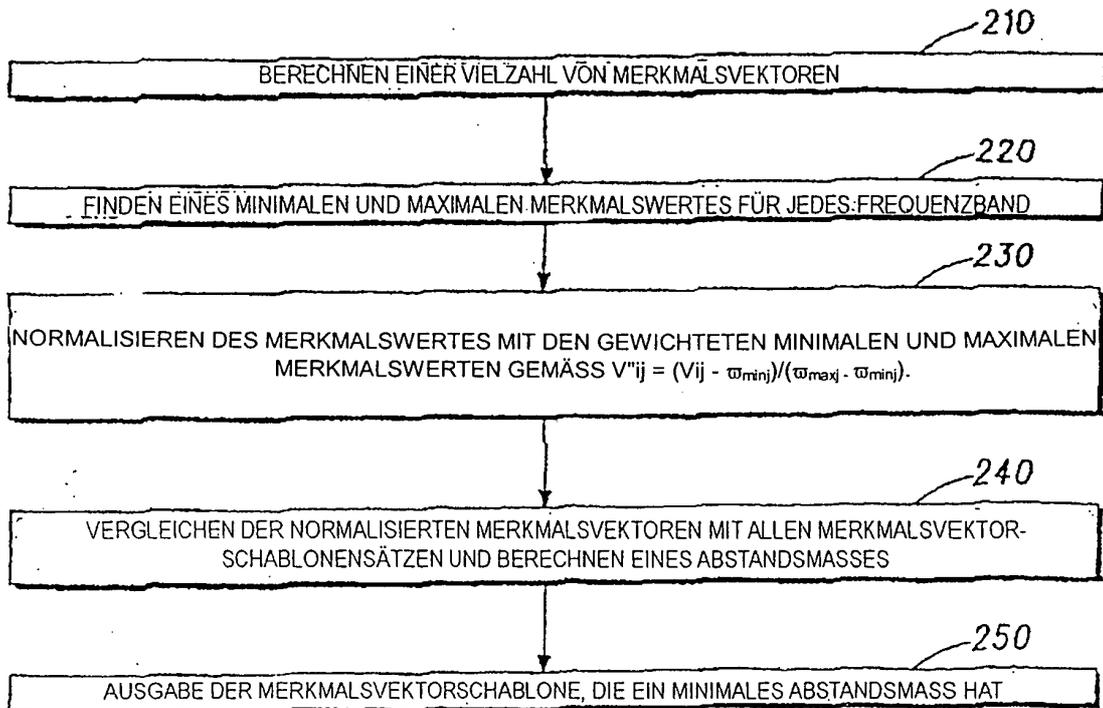


FIG.2

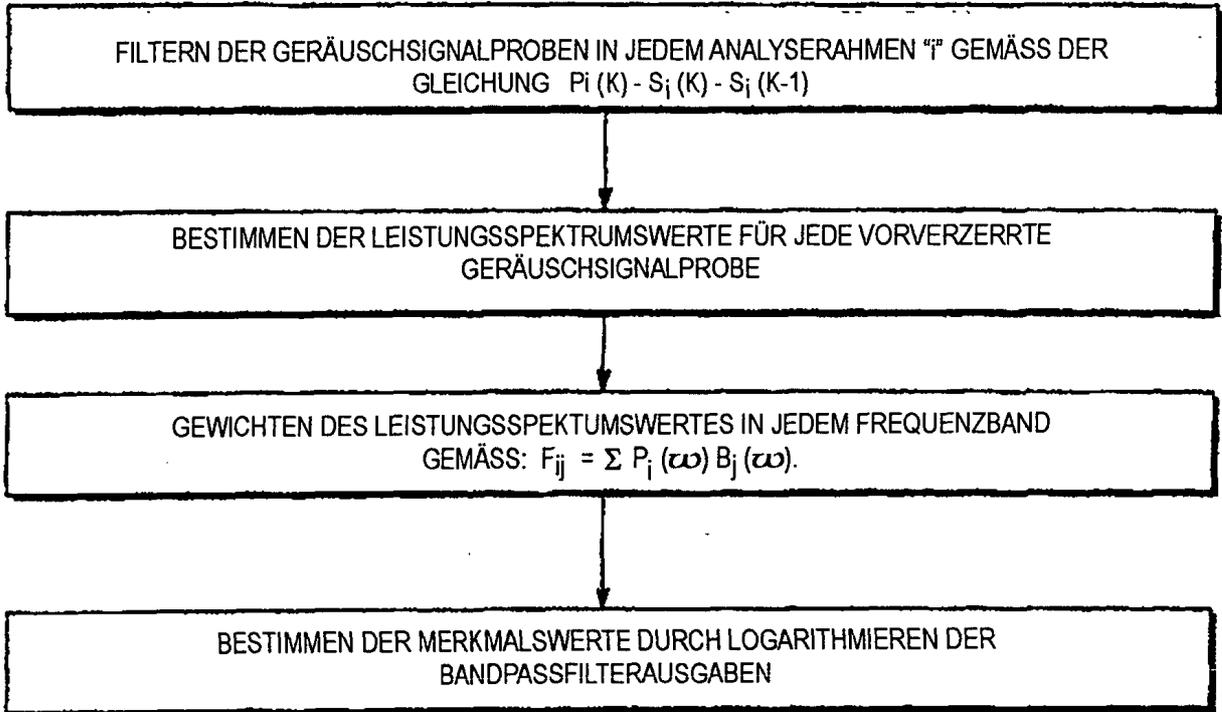


FIG.3

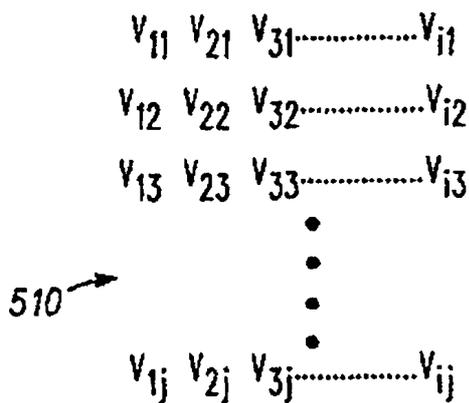


FIG.5

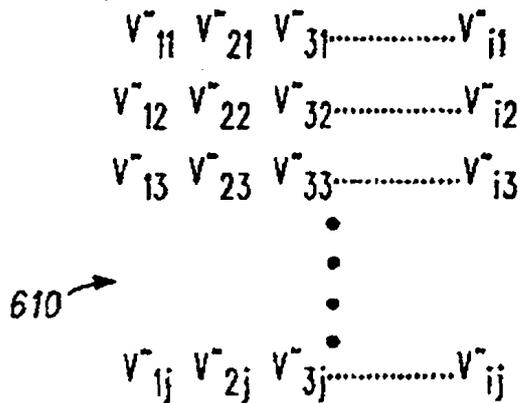


FIG.6

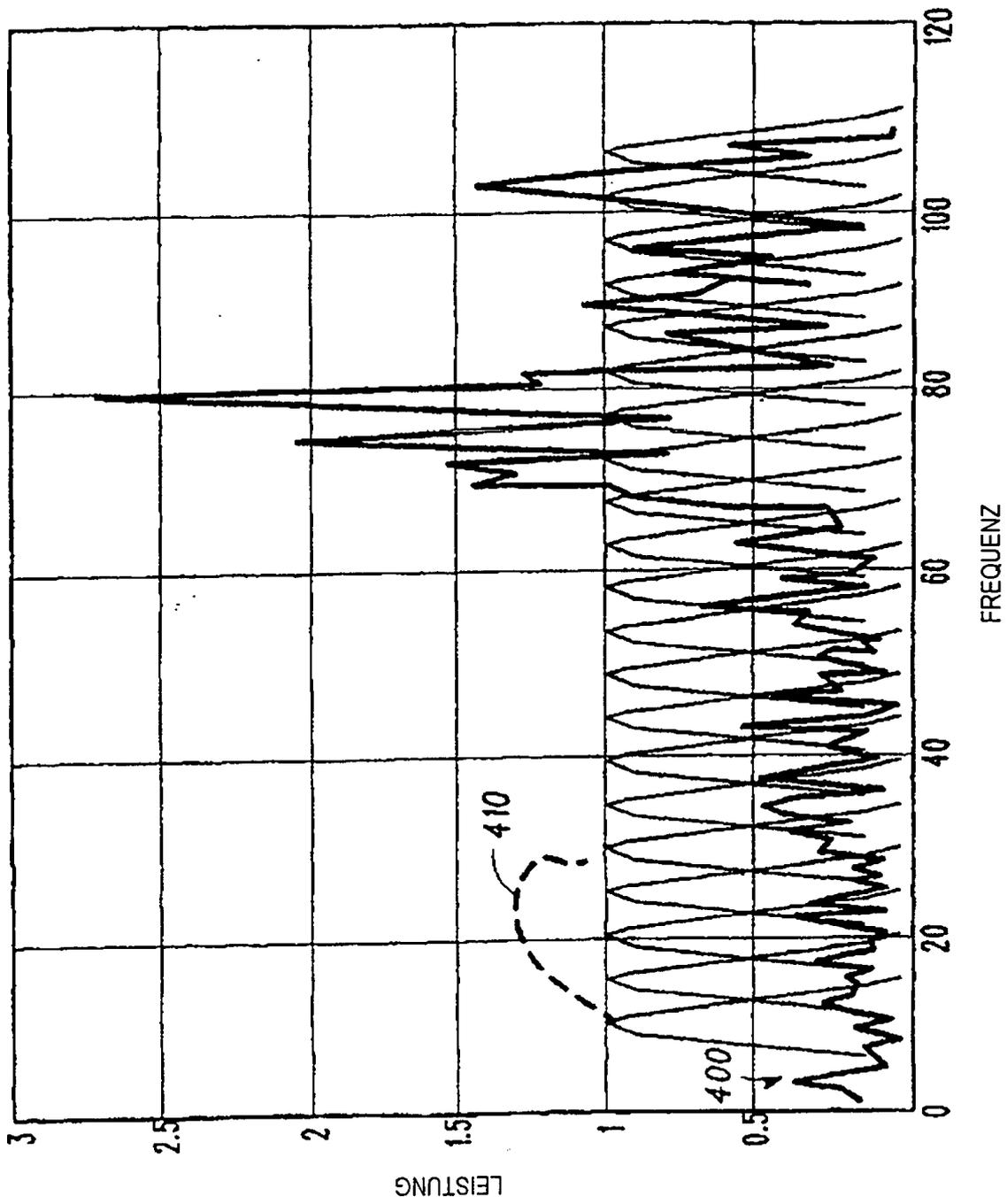


FIG.4