(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
22. März 2001 (22.03.2001)

PCT

(10) Internationale Veröffentlichungsnummer
WO 01/20598 A1

(51) Internationale Patentklassifikation[7]: G10L 21/02, G06T 5/40

(21) Internationales Aktenzeichen: PCT/AT00/00230

(22) Internationales Anmeldedatum:
28. August 2000 (28.08.2000)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
A 1559/99 10. September 1999 (10.09.1999) AT

(71) Anmelder *(für alle Bestimmungsstaaten mit Ausnahme von US)*: **SIEMENS AG ÖSTERREICH** [AT/AT]; Siemensstrasse 88-92, A-1210 Wien (AT).

(72) Erfinder; und

(75) Erfinder/Anmelder *(nur für US)*: **TSCHIRK, Wolfgang** [AT/AT]; Pelzgasse 2/13, A-1150 Wien (AT).

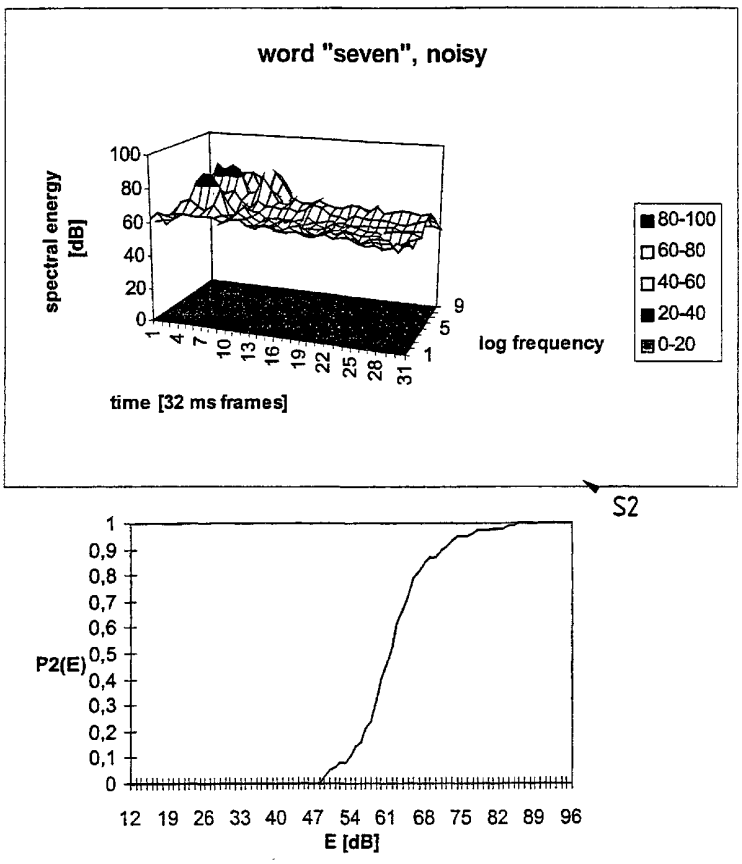
(74) Anwalt: **MATSCHNIG, Franz**; Siebensterngasse 54, A-1071 Wien (AT).

(81) Bestimmungsstaaten *(national)*: CA, CN, JP, US.

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR SUPPRESSING SPURIOUS NOISE IN A SIGNAL FIELD

(54) Bezeichnung: VERFAHREN ZUR UNTERDRÜCKUNG VON STÖRRAUSCHEN IN EINEM SIGNALFELD



(57) Abstract: The invention relates to a method for suppressing spurious noise in a signal field (S2), e.g. in a speech signal spectrum, containing a plurality of signal components which each adopt a value of a signal level and are assigned to an ordinate area (T, F). According to said method, the distribution function (P2(E)) of the signal field is first determined. As a function of the signal level, said distribution function indicates the size of the fraction of those signal components whose signal level is lower than their argument value (E). The signal level values are then modified, based on a comparison between the distribution function (P2(E)) and a reference distribution function which has been obtained from a distribution function that was determined for a set of reference models, whereby the sequence of signal components remains unchanged with regard to their energy level and signal components whose original signal levels are identical, are assigned the same modified signal levels.

[Fortsetzung auf der nächsten Seite]

WO 01/20598 A1

(84) **Bestimmungsstaaten** *(regional)*: europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht:

— *Mit internationalem Recherchenbericht.*

— *Vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eintreffen.*

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(57) Zusammenfassung: Zur Unterdrückung von Störrauschen in einem Signalfeld (S2), z.B. einem Spektrum eines Sprachsignals, enthaltend eine Vielzahl von Signalkomponenten, welche jeweils einen Wert eines Signalpegels annehmen und über einem Ordinatenbereich (T, F) auftragbar sind, wird zunächst die Verteilungsfunktion (P2(E)) des Signalfelds bestimmt. Diese gibt als Funktion des Signalpegels an, wie groß der Anteil jener Signalkomponenten ist, deren Signalpegel niedriger als ihr Argumentwert (E) ist. Sodann werden aufgrund eines Vergleiches der Verteilungsfunktion (P2(E)) mit einer Referenzverteilungsfunktion, welche aus einer für einen Satz von Referenzmustern bestimmten Verteilungsfunktion gewonnen wurde, die Signalpegelwerte des Signalfeldes modifiziert, wobei die Abfolge der Signalkomponenten hinsichtlich ihrer Energiepegel ungeändert bleibt sowie Signalkomponenten, deren ursprüngliche Signalpegel gleich sind, gleiche modifizierte Signalpegel zugewiesen werden.

VERFAHREN ZUR UNTERDRÜCKUNG VON STÖRRAUSCHEN IN EINEM SIGNALFELD

Die Erfindung betrifft ein Verfahren zur Unterdrückung von Störrauschen in einem Signalfeld enthaltend eine Vielzahl von Signalkomponenten, welche jeweils einen Wert eines Signalpegels annehmen und über einem Ordinatenbereich auftragbar sind, bei welchem aus dem Signalfeld eine Verteilungsfunktion bestimmt wird, welche als Funktion des Signalpegels zu jedem seiner möglichen Signalpegel-Argumentwerte angibt, wie groß der Anteil jener Signalkomponenten ist, deren Signalpegel niedriger als der Argumentwert ist.

Signalfelder, auf die das erfindungsgemäße Verfahren sich bezieht, werden beispielsweise in Mustererkennungssystemen zur Beschreibung der zu erkennenden Muster verwendet. Der bei der Erkennung eines Musters ablaufende Vorgang kann gewöhnlich grob in die folgenden Schritte aufgeteilt werden: Erfassung des Musters, Vorverarbeitung und Klassifizierung.

Der erste Schritt, die Mustererfassung, dient der Umwandlung des originalen Musters, z.B. eine gesprochene Äußerung eines Benutzers oder ein mit Text beschriebenes Schriftstück, in ein für die Verarbeitung geeignetes Format, z.B. in Form eines elektronischen Signals, das analog oder digital codiert sein kann, oder einer Datei vorgegebenen Formats. Hierher gehört auch die Umwandlung eines Signal/Dateiformats, z.B. einer Rasterbildaufnahme, in ein für die weitere Verarbeitung geeignetes Format. Im Falle einer Spracherkennung beispielsweise wird die vom Benutzer gesprochene Äußerung über eine akustische Eingabe, wie z.B. ein Mikrophon, aufgenommen, gegebenenfalls vorverstärkt und in ein elektrisches Sprachsignal in analoger oder digitalisierter Form umgesetzt.

Das so erfasste Muster wird der Vorverarbeitung zugeführt, die eine Reduktion der zu verarbeitenden Daten sowie eine bessere Unterscheidbarkeit der zu bestimmenden Muster erreicht. Ergebnis der Vorverarbeitung ist ein Signalfeld, im Beispiel der Spracherkennung ein Spektrum der Äußerung, das dem Klassifizierungssystem zugeführt werden kann. Häufig ist ein wesentlicher Schritt der Vorverarbeitung eine Signalanalyse des Mustersignals, z.B. kann für das elektrische Sprachsignal der Benutzer-Äußerung eine Signalanalyse in Form einer Aufteilung in Zeitrahmen (Diskretisierung) und einer nachfolgenden, jeweils innerhalb eines Zeitrahmens durchgeführten Fourier-Transformation mit Zerlegung in Frequenzbänder erfolgen, aus der ein Zeit-Frequenz-Spektrum gewonnen wird. Damit ist zugleich eine - im allgemeinen beträchtliche - Datenreduktion verbunden. Ein weiterer, unter Umständen wesentlicher Schritt der Vorverarbeitung ist die Verringerung von Störrauschen in dem Mustersignal bzw. dem daraus gewonnen Signalfeld.

Das Signalfeld umfasst eine Vielzahl von Signalkomponenten, die jeweils einen eigenen, hier als Signalpegel bezeichneten Wert gleichen Typs annehmen. Die Signalkomponenten sind naturgemäß innerhalb des Signalfelds geordnet, wobei diese Ordnung mit Hilfe eines oder mehrerer Ordinatenparameter ausgedrückt ist. Beispielsweise besteht ein als Zeit-Frequenz-Spektrum realisiertes Signalfeld aus vielen Spektralkomponenten, die jeweils einen eigenen Energiepegel annehmen; die Spektralkomponenten sind nach Zeitrahmen und Frequenzband geordnet. Jeder Signalkomponente kann somit in dem Ordinatenbereich, über den sich das Signalfeld erstreckt, ein eigener Bereichselement des Ordinatenbereichs zugeordnet werden, sodass die Bereichselemente insgesamt den Ordinatenbereich des Signalfeld abdecken. In Abhängigkeit von der Anzahl der Ordinatenparameter kann der Ordinatenbereich ein-, zwei- oder mehrdimensional sein; dementsprechend sind die Bereichselemente Linien-, Flächen- oder (n-dimensionale) Volumselemente.

Das durch die Vorverarbeitung erhaltene Signalfeld wird dem Klassifizierungssystem zugeführt. Dieses ermittelt, zu welcher Erkennungsklasse - d.i. im Falle der Spracherkennung ein Wort eines vorgegebenen Wortschatzes oder eine Wortkette - eine Übereinstimmung gegeben ist. Das Erkennungsresultat wird dann Ausgabe zugeführt, beispielsweise auf einer Anzeige, oder zur weiteren Verarbeitung genutzt, z.B. bei einer Befehlseingabe einer sprachorientierten Einrichtung.

Die Ausführung einer Mustererkennung wird oftmals durch Störrauschen erschwert, das die zu erkennenden Muster überlagert. Beispielsweise kann die Leistungsfähigkeit eines Spracherkennungssystems durch akustischen Hintergrundlärm stark herabgesetzt oder ganz vereitelt werden.

Bei bekannten Verfahren zur Rauschunterdrückung wird in der Vorverarbeitung eine Abschätzung der dem Signal unterliegenden Rauschparameter durchgeführt und aufgrund dieser Abschätzung ein Referenzrauschsignal abgezogen wird. Derartige Verfahren der spektrale Subtraktion für Sprachsignale werden von S. V. Vaseghi und B. P. Milner in 'Noise Compensation Models for Hidden Markov Model Speech Recognition in Adverse Environments', IEEE Transactions on Speech and Audio Processing, Vol. 5, No. 1, Januar 1997, S. 11-21 beschrieben. Hierbei wird von dem Energiepegel E jeweils einer Spektralkomponente des Spektrums die entsprechende Komponente eines Referenzrauschsignals E_r gemäß dem Ausdruck

$$E' = s_s(E, E_r) = (E^b - \alpha E_r^b)^{1/b}$$

„subtrahiert". Das Referenzrauschsignal E_r wird aufgrund vorgegebener oder abgeschätzter Rauschparameter simuliert. Die Subtraktion der Energiepegel kann hierbei z.B. in Bezug auf

die linearen Energiepegel durchgeführt werden oder „konvolutiv" im logarithmischen Bereich, d.h. in der genannten Formel stehen anstelle der Energiepegel E, E_r, E' die entsprechenden Logarithmen log E, etc.

Der Subtraktionsansatz hat jedoch den Mangel, dass die zur Beschreibung des Rauschens notwendigen Parameter nicht mit der erforderlichen Genauigkeit und Vollständigkeit bekannt sein können. Beispielsweise ist für eine korrekte Rauschkompensation nicht nur die Kenntnis der Rauschamplituden, sondern auch der Phasenbeziehungen erforderlich, was - wenn überhaupt - nur mit sehr großem Aufwand möglich ist. Störungen, die keine additive oder konvolutive Überlagerung darstellen, wie z.B. Mischformen aus additiven und konvolutiven Störungen, sind noch schwieriger zu behandeln.

Die EP 0 062 519 A1 lehrt die Beseitigung von Störungen in Radarsignalen, wobei die Verteilung der Störungen bekannt, wenn auch beliebig ist, im Gegensatz zu vorher bekannten Verfahren, die eine Rayleigh- oder Weibull-verteilte Störung verlangen. Die Kenntnis der Verteilung oder zumindest der zugehörenden Wahrscheinlichkeitsdichte, aus der man sie ableiten kann, ist notwendige Voraussetzung für die Anwendung des Verfahrens dieses Dokuments. Ohne Kenntnis einer solchen Verteilung ist eine Störungsbeseitigung nach diesem Verfahren somit nicht durchführbar.

Die EP 0 548 527 A2 lehrt ein Verfahren zur Erzeugung einer Transformation der Pegelskala eines digitalen radiographischen Bildes, z.B. Röntgenbildes, in welchem eine kumulative Verteilungsfunktion des Bildes verwendet wird, um die Pegelverteilung des Bildes dahingehend zu modifizieren, dass sie im interessierenden Bereich im wesentlichen linear ist. Die diesem Verfahren zugrundeliegende Aufgabenstellung, nämlich eine Darstellung des Bildes in einer für die weitere Untersuchung durch Betrachtung des Bildes geeigneten Form, unterscheidet sich freilich wesentlich von jener der Erfindung.

Die EP 0 720 358 A2 betrifft die Kompression von Videosignaldaten. Dabei wird die Pegelverteilung eines Bildes so modifiziert, dass jedem Eingangspegelbereich ein um so größerer Ausgangspegelbereich zugeordnet wird, je mehr Eingangspegel in ersteren Bereich fallen, wobei der gesamte Ausgangspegelbereich begrenzt ist. Auch in diesem Fall ist die Aufgabenstellung, nämlich eine gleichmäßigere Signalkompression, von jener der Erfindung wesentlich verschieden. Dem entsprechend wird bei der Kompression nach dieser Schrift eine Zielverteilung nicht angestrebt; vielmehr verwendet die Kompressionsvorschrift lediglich aus dem Eingangssignal abgeleitete Parameter.

Aus keinen der genannten Dokumente geht die Verwendung einer aus Trainings- oder Referenzdaten gewonnenen Referenzverteilungsfunktion hervor.
Es ist daher Aufgabe der Erfindung, ein Verfahren zur Rauschunterdrückung aufzuzeigen, das die Beeinträchtigung des Signalfelds durch das Störrauschen hinsichtlich der nachfolgenden Auswertung, insbesondere einer Klassifizierung, zuverlässig verringert. Weiters soll die Rauschunterdrückung ohne nähere Kenntnis der Eigenschaften des Rauschens und ohne eine Simulation eines Hintergrundrauschens durchführbar sein.

Die Aufgabe wird von einem Verfahren der eingangs genannten Art gelöst, bei welchem erfindungsgemäß aus dem Signalfeld eine Verteilungsfunktion bestimmt wird, welche als Funktion des Signalpegels zu jedem seiner möglichen Signalpegel-Argumentwerte angibt, wie groß der Anteil jener Signalkomponenten ist, deren Signalpegel niedriger als der Argumentwert ist, und sodann aufgrund eines Vergleiches der Verteilungsfunktion mit einer vorbestimmten Referenzverteilungsfunktion die Signalpegelwerte des Signalfeldes modifiziert werden, wobei die Abfolge der Signalkomponenten hinsichtlich ihrer Energiepegel ungeändert bleibt sowie Signalkomponenten, deren ursprüngliche Signalpegel gleich sind, gleiche modifizierte Signalpegel zugewiesen werden, wobei als Referenzverteilungsfunktion eine aus einer Verteilungsfunktion, die für einen Satz von Referenzmustern bestimmt worden ist, gewonnene Funktion verwendet wird.

Diese Lösung ermöglicht eine Rauschunterdrückung sowohl für additiven bzw. konvolutiven Rauschhintergrund als auch für Mischformen oder noch kompliziertere Störungen. Durch das erfindungsgemäße Verfahren kann die Auswirkung der Störung auf die Signalparameter des Signalfelds beträchtlich reduziert werden, auch ohne nähere Kenntnis von Rauschparametern.

Die Forderung, dass die Abfolge der Signalkomponenten hinsichtlich ihrer Energiepegel ungeändert bleibt, bedeutet, dass für jedes (beliebige) Paar von Signalkomponenten, für welche der ursprüngliche Pegel der ersten Komponente kleiner als jener der zweiten ist, nach der Zuweisung modifizierter Pegel zu den Signalkomponenten der modifizierte Pegel der ersten Komponente nicht größer (also gleich oder kleiner) als der modifizierte Pegel der zweiten Komponente ist.

Es sei darauf hingewiesen, dass sich aus den oben genannten Schriften keinerlei Hinweise entnehmen lassen, dass eine Modifikation anhand einer Referenzverteilungsfunktion ohne Berücksichtigung der Art des Störrauschens erfolgreich sein könnte.

Der für das erfindungsgemäße Verfahren wesentliche Parameter, die Referenzverteilungsfunktion, kann im vorhinein z.B. mit Hilfe von Versuchen bestimmt werden. Wenn ein Trainings- oder Vergleichssatz von Mustern vorliegt, können diese oder eine ausgewählter Teil dieser Muster zur Erzeugung der Referenzverteilungsfunktion dienen. Vorteilhafterweise kann dann als Referenzverteilungsfunktion eine aus einer Verteilungsfunktion, die für einen Satz von Referenzmustern bestimmt worden ist, gewonnene Funktion verwendet werden. Dabei kann die Verteilungsfunktion des Referenzmustersatzes selbst als Referenzverteilungsfunktion genutzt werden, oder eine aus ihr, z.B. durch Vereinfachung des Kurvenverlaufs, gewonnene Funktion des Pegels.

Günstigerweise erfolgt die Modifizierung der Signalpegelwerte dadurch, dass ausgehend von einer Aufteilung des Wertebereichs der Signalpegel in eine Anzahl von Pegelbereichen für jeden Pegelbereich

- zu einem diesen Pegelbereich repräsentierenden, ersten Pegel unter Anwendung der Verteilungsfunktion und des Werts der Referenzverteilungsfunktion an dem ersten Pegel ein zweiter Pegel ausgewählt wird, für welchen der Wert der Verteilungsfunktion dem genannten Wert der Referenzverteilungsfunktion möglichst nahe kommt, und
- jenen Signalkomponenten, deren Signalpegel zwischen dem ersten und dem zweiten Pegel fällt, der Wert des ersten Pegels zugewiesen wird.

Dies erlaubt eine möglichst weitgehende Anpassung des Signals an die Referenzverteilungsfunktion. Im einfachsten Falle der Aufteilung des Signalpegel-Wertebereichs in Pegelbereiche wird für jeden auftretenden Signalpegel ein eigener Bereich zugeordnet, sodass jeder Pegelbereich mit dem zugehörenden Signalpegel identifiziert werden kann.

Des weiteren wird eine besonders zweckmäßige Realisierung der Erfindung für ein als zeit- und/oder frequenzabhängiges Spektrum eines akustischen Signals realisiertes Signalfeld ausgeführt.

Die Erfindung wird im folgenden anhand eines Ausführungsbeispiels erläutert, das die Spracherkennung eines gesprochenen Wortes in einem Kraftfahrzeugwagen betrifft. Dabei werden die beigefügten Figuren herangezogen, welche zeigen:

Fig. 1 ein Spektrogramm einer Äußerung unter geräuschfreien Bedingungen;

Fig. 2 die Energieverteilungsfunktion zu dem Spektrogramm der Fig. 1;

Fig. 3 und 4 ein Spektrogramm und die zugehörende Energieverteilungsfunktion einer Äußerung mit Geräuschhintergrund;

Fig. 5 und 6 ein Spektrogramm und die zugehörende Energieverteilungsfunktion, welche sich durch spektrale Subtraktion aus dem Spektrogramm der Fig. 3 ergeben;

Fig. 7 eine Referenzverteilungsfunktion zur Anwendung der Erfindung;

Fig. 8 und 9 ein Spektrogramm und die zugehörende Energieverteilungsfunktion, welche sich aus dem Spektrogramm der Fig. 3 mittels der erfindungsgemäßen Rauschreduktion anhand der Referenzverteilungsfunktion der Fig. 7 ergeben.

Sprachsignale, welche vor einem Geräuschhintergrund, wie z.B. jenem im Inneren eines Kraftfahrzeugwagens in Betrieb, gesprochen werden, werden durch Geräusche beeinträchtigt, die von verschiedenen Quellen, z.B. dem Fahrzeugmotor, anderen Fahrzeugen, Wind usw., stammen können und oftmals eine Mischung von Schallkomponenten hoher Energie mit nicht vorhersehbarer Statistik hinsichtlich ihres Zeitablaufs und ihrer Frequenz darstellen. Die Leistungsfähigkeit von Spracherkennungssystemen nimmt daher schnell ab, wenn der Geräuschhintergrund zunimmt, beispielsweise weil die Fahrzeuggeschwindigkeit größer wird. Das im folgenden dargestellte Ausführungsbeispiel der Erfindung betrifft die Erkennung der englischen Wörter 'zero', 'one', 'two', usw. bis 'nine' für die Ziffern 0 bis 9 mittels eines Spracherkennungssystems in einem Wagen vom KFZ-Kleinwagentyp.

Fig. 1 zeigt ein Spektrogramm S1 eines Spektrums zu einer Äußerung des englischen Wortes 'seven', gesprochen von einem männlichen Sprecher in dem Wagen unter geräuschfreien Bedingungen.

In den in dem Ausführungsbeispiel behandelten Spektren erfasst die Zeitachse einen Zeitraum von 0.992 s, die in 31 Rahmen T gleicher Zeitdauer (sogenannte 'frames') aufgeteilt ist. Der Frequenzbereich erstreckt sich von f = 200 Hz bis 3.4 kHz und ist in 9 Bänder F mit ungefähr logarithmisch abgestufter Bandbreite und -abstand aufgeteilt. Die spektrale Energie ist in allen Figuren logarithmisch als Energiepegel E, mit der Einheit dB und bezogen auf einen allen Figuren gemeinsamen Grundpegel, dargestellt.

Spektren dieser Art wurden in Spracherkennungsversuchen der Anmelderin für Äußerungen über den genannten Wortschatz verwendet. In dem verwendeten Spracherkennungssystem erfolgt nach einer Vorverarbeitung des zu erkennenden Äußerung mittels einer Rauschunterdrückung wie weiter unten näher erläutert eine Klassifizierung, bei welcher ein geschichtetes neuronales Netzwerk, welches mit einem Trainingswortschatz trainiert worden war, als Mustererkennungssystem dient. Für den Trainingswortschatz wurde der Wortschatz von einer Anzahl von Sprechern - vorteilhafterweise sowohl männliche als auch weibliche Personen - in einer Umgebung, die der Sprechumgebung des Wagens entspricht, gesprochen, und zwar für jedes Wort jeweils mehrere Male unter rauschfreien Bedingungen des Rauschhintergrunds (Ruhe des Wagens).

Fig. 2 zeigt die Energieverteilungsfunktion P1(E) zu dem in Fig. 1 dargestellten Spektrum S1 . Eine einem Spektrum S zugeordnete Energieverteilungsfunktion P(E) gibt als Funktion des Energiepegels E an, wie viele der spektralen Komponenten S(T,F) des betreffenden Spektrums S einen Energiepegel aufweisen, der niedriger als der angegebene Energiepegel E ist, wobei diese Zahl als Wert zwischen 0 und 1 bezogen auf die Gesamtzahl der spektralen Komponenten ausgedrückt ist. Beispielsweise hat die Energieverteilungsfunktion P1 bei 48 dB den Wert 0.6, denn 60 % der Energiepegel des Spektrums S1 liegen unter 48 dB. Eine große (kleine) Steigung in der Energieverteilungsfunktion P(E) entspricht einem Energiepegel, dessen Wert in einer großen (kleinen) Anzahl von Komponenten des zugehörenden Spektrums S auftritt. Eine Energieverteilungsfunktion kann auch für eine Vielzahl von Spektren bestimmt werden und gibt dann den Anteil der Komponenten sämtlicher Spektren mit Energiepegel unter dem angegeben Pegel E, geteilt durch die Gesamtzahl der Komponenten aller dieser Spektren, an.

Fig. 3 zeigt das Spektrogramm S2 zu einer Äußerung des Wortes von demselben Sprecher bei einer Wagengeschwindigkeit von 113 km/h (70 mph). Wie aus dem Vergleich der Spektrogramme S1 und S2 (Fig. 1 bzw. 3) ersichtlich, bleiben lediglich die Sprachanteile hoher Energie wenig beeinträchtigt, während die übrigen Anteile von den Geräuschen maskiert sind. Der Hintergrund-Energiepegel steigt von ungefähr 25 dB auf ungefähr 65 dB, die Spitzen der Äußerung sind bei 85 dB, die Sprachanteile unterhalb 70 dB gehen im Geräuschhintergrund unter. Die zugehörende Energieverteilungsfunktion P2(E) ist in Fig. 4 dargestellt.

Die Energieverteilungsfunktionen P1 und P2 (Fig. 2 bzw. 4) zeigen, dass die spektrale Verteilung des rauschfreien Signals S1 deutlich verschieden von jener des geräuschbehafteten Signals S2 ist, in dem die Hintergrundenergie um ungefähr 40 dB höher liegen als im Falle des rauschfreien Signals.

Mittels der eingangs erwähnten spektralen Subtraktion nach S. V. Vaseghi und B. P. Milner ist eine Rauschreduktion des verrauschten Signals erreichbar. Entsprechend dem weiter oben Gesagten, wird das Spektrum S unter Verwendung eines Referenzrauschsignals S_r dadurch transformiert, dass in jeder Spektralkomponente S(T,F) die jeweils entsprechende Komponente $S_r(T,F)$ des Referenzrauschens gemäß dem Ausdruck

$$S'(T,F) = E0 = s_s(E, E_r) = (E^b - \alpha E_r^b)^{1/b} \text{ , wobei } E = S(T,F) \text{ und } E_r = S_r(T,F)$$

„subtrahiert" wird. Die Rauschreduktion nach der spektralen Subtraktion wurde im Rahmen der weiter unten beschriebenen Versuche der Anmelderin für das Spektrum S2 durchge-

führt. In Fig. 5 und 6 sind das Spektrum S3 = s_S(S2, S_r), das sich bei der Anwendung der spektralen Subtraktion auf das Spektrogramm S2 ergibt, und die zugehörende Energieverteilungsfunktion P3 dargestellt; dabei wurden jene Parameter b und α verwendet, bei denen die Ergebnisse von durchgeführten Spracherkennungstests für verschiedene Parameter b und α am besten waren, sowie ein aus der Aufnahme der Äußerung S2 gewonnenes Referenzrauschen S_r. Wie aus Fig. 5 und 6 ersichtlich ist, ist das Hintergrundrauschen ist um ca. 10 dB niedriger als im unbehandelten Signal S2, jedoch ist ein beträchtlicher Anteil der Sprachanteile niedriger Energie immer noch vom restlichen Rauschen verdeckt. Daher verbessert sich die Erfolgsquote bei der Spracherkennung nur geringfügig.

Da das als Referenzrauschsignal S_r verwendete Signal nur statistisch mit dem Rauschen übereinstimmt, welches als Hintergrund des verrauschten Signals S2 vorliegt, erzielt die spektrale Subtraktion eine Reduktion des Rauschpegels nur an einzelnen Komponenten des sich ergebenden Spektrums S3. Denn in Abhängigkeit von der relativen Phasenlage des Referenzrauschens und des tatsächlichen Hintergrunds kommt es nur für einen Teil der Komponenten des Spektrums zu einer Auslöschung des Rauschanteils der betreffenden Komponente, in anderen Komponenten bleibt der Pegel ungefähr gleich, in manchen ergibt sich sogar eine Verstärkung (wenngleich deren Auswirkung aufgrund der logarithmischen Darstellung der Energiepegel gemildert ist). Dies ist in Fig. 5 besonders an den Niedrigpegel-Anteilen ca. ab Zeitrahmen 20 zu erkennen.

Gemäß der Erfindung erfolgt die Rauschunterdrückung für das vorliegende Sprachsignal S2 unter Verwendung einer vorgegebenen „Vorlagefunktion", nämlich einer als Referenz dienenden Energieverteilungsfunktion. Vorteilhafterweise geschieht dies derart, dass die Pegel der Spektralkomponenten des Sprachsignal-Spektrums S2 an die Vorlagefunktion angepasst werden. Die Energieverteilungsfunktion des sich ergebenden Spektrums stimmt dann im wesentlichen mit der Vorlagefunktion überein.

Idealerweise würde als Vorlagefunktion die Energieverteilungsfunktion der Summe jener Spektren verwendet werden, welche beim Training des Spracherkennungssystems für das betreffende Wort (hier 'seven') verwendet werden; da das zu erkennende Wort dem Spracherkennungssystem naturgemäß nicht im vorhinein bekannt ist, ist dies nicht möglich. Es wird stattdessen eine Energieverteilungsfunktion als Vorlagefunktion gewählt, welche in Bezug auf die Gesamtheit der Worte des zu erkennenden Wortschatzes zweckmäßig ist. Beispielsweise kann jene Energieverteilungsfunktion als Vorlagefunktion P0 verwendet werden, welche aus den Spektren des gesamten Trainingswortschatzes abgeleitet wurde.

Die erfindungsgemäße Rauschunterdrückung durch Anpassung der Pegel an eine Vorlagefunktion erfolgt derart, dass Spektralkomponenten, deren Pegel $E = S(T,F)$ ursprünglich gleich ist, auch nach der Anpassung einen gemeinsamen Pegel $E0 = S'(T,F)$ aufweisen, d.h. für alle Spektralkomponenten gilt die Anpassungsbedingung

$$S'(T_1,F_1) = S'(T_2,F_2) \quad \text{wenn} \quad S(T_1,F_1) = S(T_2,F_2)\,. \qquad (1)$$

De weiteren soll die Abfolge der Komponenten hinsichtlich ihrer Energiepegel nicht geändert werden, d.h.

$$S'(T_1,F_1) \le S'(T_2,F_2) \quad \text{wenn} \quad S(T_1,F_1) < S(T_2,F_2)\,; \qquad (2)$$

diese Monotoniebedingung bewahrt bei der Rauschunterdrückung des Spektrums S in ein modifiziertes Spektrum S' die Strukturen des Spektrums zumindest in qualitativer Hinsicht.

Die Rauschunterdrückung kann als Konsequenz der Anpassungsbedingung (1) durch eine Anpassungsfunktion R(E) vollständig beschrieben werden, die jedem ursprünglichen Pegel E einen modifizierten Pegel $E0 = R(E)$ zuordnet, auf welchen jene Spektralkomponenten gesenkt (oder gehoben) werden, die ursprünglich den Pegel E aufwiesen. Die Anpassungsfunktion ist wegen der Monotoniebedingung (2) monoton, d.h. $R(E_1) \le R(E_2)$ wenn $E_1 < E_2$. Erfindungsgemäß erfolgt dies Anpassung des Spektrums derart, dass für die zugeordnete Energieverteilungsfunktion gilt $P0(E0) = P(E)$. Daher ist die Anpassungsfunktion R(E) eindeutig durch den Vergleich der Energieverteilungsfunktion P2 des vorliegenden Signals mit der Vorlagefunktion P0 bestimmt. Da die Energieverteilungsfunktionen P,P0 gleichfalls monoton wachsende Funktionen sind, kann formal daraus die Anpassungsfunktion mittels Umkehrung der Vorlagefunktion P0 ermittelt werden.

Tabelle 1 zeigt einen beispielhaften Programm-Pseudocode, durch den die erfindungsgemäße Anpassung eines Spektrums erfolgt. Das anzupassende Spektrum S ist hierbei in der Feldvariablen `S` gespeichert, das über die Intervalle `Tmin..Tmax` sowie `Fmin..Fmax` des Zeit-Frequenz-Raumes definiert ist. Die Energiepegel des Spektrums können diskrete Werte in dem Wertebereich zwischen den Energiepegeln `Emin` und `Emax` annehmen. In der Feldvariablen `P0` ist eine Referenz-Energieverteilungsfunktion als Vorlagefunktion vorgegeben. Die Energieverteilungsfunktionen sind als Felder über das genannte Intervall `Emin..Emax` definiert.

Zunächst (ab der Marke PS/S) wird die zugehörende Energieverteilungsfunktion ermittelt und in der Feldvariablen `PS` abgelegt. Hierzu wird für jede Komponente `S[T,F]` des Spektrums der Pegelwert ermittelt, und sämtliche Komponenten der Energieverteilungsfunktion

```
{ PS/S }
for E = Emin to Emax :
  PS[E] = 0;
end for;
for T = Tmin to Tmax : for F = Fmin to Fmax :
  for E = S[T,F] to Emax :
     inc(PS[E]);
  end for;
end for; end for;

{ RED/S }
for E0 = Emin to Emax :
  if P0[E0] > PS[E0] :

    dE = 0;
    while E0+dE<=Emax
          and  abs( P0[E0]-PS[E0+dE] )<=abs( P0[E0]-PS[E0+dE-1] ) :
      inc(dE);
    end while;
    dec(dE);

    if dE > 0 :
      for T = Tmin to Tmax : for F = Fmin to Fmax :
        if S[T,F] > E0 and S[T,F] <= E0+dE :
          S[T,F] = E0;
        end if;
      end for; end for;
    end if;

  end if;
end for;
```

Tabelle 1

PS, deren zugeordneter Energiepegel über diesem Pegelwert liegt, werden inkrementiert. Hierbei bezeichnet `inc` die Inkrementierfunktion.

Sodann (ab der Marke RED/S) wird in einer for-Schleife für jeden der diskreten Werte `E0`, sofern an diesem Pegel die Energieverteilungsfunktion `PS[E0]` kleiner als die Vorlagefunktion `P0[E0]` ist, die folgenden Schritte ausgeführt: Es wird zunächst ein dem Pegelwert `E0` zugeordneter Energiepegel `E0+dE` bestimmt. Dies geschieht dadurch, dass der Abstand `dE` dieser Pegel ausgehend von dem Wert 0 solange inkrementiert wird (while-Schleife), bis der Wert der Energieverteilungsfunktion am zugeordneten Pegel `PS[E0+dE]` dem Wert der Vorlagefunktion am gegebenen Pegelwert `P0[E0]` am nächsten kommt. Hierzu wird die Funktion `abs` zur Ermittlung des Absolutbetrages verwendet. Der nach der while-Schleife stattfindende Dekrementierschritt `dec(dE)` dient der Korrektur auf jenen Wert, für welchen die genannte Bedingung tatsächlich zutrifft. Nun stellt der Pegelwert `E0` den modifizierten Pegel zu dem Energiepegel `E0+dE` dar. Sodann wird geprüft, ob der Pegelabstand `dE` positiv

(größer als 0) ist; in diesem Fall werden sämtliche Komponenten S[T,F] des Spektrums, deren Energiepegel in das Intervall zwischen E0 und E0+dE fällt, auf den Energiepegel E0 gestellt. Nach dem letzten Durchlauf der äußeren for-Schleife enthält das Feld S das erfindungsgemäß rauschunterdrückte Spektrum S'.

Fig. 7 zeigt die in dem Ausführungsbeispiel verwendete Vorlagefunktion P0(E0), nämlich die Energieverteilungsfunktion für den oben genannten Trainingswortschatz, d.s. die englischen Zahlwörter 'zero' bis 'nine'. Für die verrauschte Äußerung S2 ergibt die erfindungsgemäße Rauschunterdrückung mit Hilfe der genannten Vorlagefunktion P0 das als Spektrogramm S4 in Fig. 8 gezeigte Spektrum; die zugehörende Energieverteilungsfunktion P4 ist in Fig. 9 wiedergegeben.

Zur Verringerung des Aufwands bei der Durchführung des erfindungsgemäßen Verfahrens kann jeweils ein Pegelbereich des ursprünglichen Spektrums derart gemeinsam behandelt werden, dass den zugehörenden Spektralkomponenten ein einheitlicher modifizierter Pegel zugewiesen wird. Dieser modifizierte Pegel wird in Bezug auf einen respräsentativen Pegelwert des betreffenden Pegelbereichs, z.B. den Mittelwert des Pegelbereichs oder den Median der Pegel über die in den in den Pegelbereich fallenden Komponenten, wie oben beschrieben bestimmt, beispielsweise mittels der Anpassungsfunktion.

Bei von der Anmelderin durchgeführten ersten Spracherkennungsversuchen mit dem oben beschriebenen Spracherkennungssystem wurde das erfindungsgemäße Verfahren getestet und zugleich mit dem Verfahren der spektralen Subtraktion verglichen. Die zu erkennenden Äußerungen wurden unter verschiedenen Bedingungen des Rauschhintergrunds gesprochen, nämlich Fahrt bei 80 km/h (50 mph) und bei 113 km/h (70 mph). Es wurden hierbei die Ereignisse gezählt, bei denen das Spracherkennungssystem die Äußerung falsch erkannt hat, wobei nur Substitutionsfehler berücksichtigt wurden. Bei einer Kontrollreihe, in der die Signale ohne Rauschreduktion der Mustererkennung zugeführt wurden, wurden 30 % der Äußerungen falsch erkannt. Bei Einsatz der spektralen Subtraktion als Rauschreduktionsverfahren ging der Anteil der fehlerhaften Erkennungen auf 23.3 % zurück. Mit dem erfindungsgemäßen Verfahren verringerte sich der Fehleranteil auf 13.3 %, also eine Reduktion der Fehlerrate um fast die Hälfte im Vergleich zum bekannten Verfahren.

Das erfindungsgemäße Verfahren eignet sich insbesondere zur Unterdrückung überlagernder Störungen, welche die Monotonierelation der Spektralkomponenten der Äußerung nicht oder nur geringfügig stören. Zu derartigen Störungen gehören z.B. weißes Rauschen, eine lineare oder nichtlineare Verstärkung oder Abschwächung des gesamten Spektrums sowie verschiedene Phänomene des Lombard-Effekts, der bekanntermaßen eine Änderung der

Stimme und der Aussprache in Abhängigkeit von dem psychischen Zustand des Sprechers, z.B. Stress, beschreibt.

In dem Spektrogramm S4 der Fig. 8 ist um Zeitrahmen 16 bei den oberen Frequenzbändern ein Artefakt erkennbar, welches in der eigentlichen Äußerung (Fig. 1) nicht enthalten ist und von dem erfindungsgemäßen Verfahren nicht beseitigt wurde. Derartige Artefakte können in den meisten Fällen z.B. mit Hilfe einer der Rauschunterdrückung nachgeschalteten Medianfilterung eliminiert werden.

Das erfindungsgemäße Verfahren der Rauschunterdrückung verändert das zu verarbeitende Signal auch bei Abwesenheit von Rauschen, da die Vorlagefunktion P0 im allgemeinen von der Energieverteilungsfunktion der ungestörten Äußerung verschieden ist. Hierdurch kann unter Umständen eine Quelle für Erkennungsfehler im rauschfreien Fall entstehen. Um dies zu vermeiden, kann beispielsweise das Training des Spracherkennungssystems mit Hilfe von Spektren durchgeführt werden, die bereits mit dem erfindungsgemäßen Verfahren an die verwendete Vorlagefunktion angepasst worden sind. Der Trainingswortschatz kann diese Spektren anstelle von oder gemeinsam mit den ursprünglichen Spektren enthalten.

Ein anderer Ansatz besteht darin, das erfindungsgemäße Verfahren nur dann einzusetzen, wenn das Vorliegen von Rauschen festgestellt wird, z.B. im Zeitraum kurz vor der Äußerung; anderenfalls wird das Sprachsignal der Spracherkennung ohne Rauschunterdrückung zugeführt. Dieser Ansatz benötigt keine Abschätzung des Rauschens, die über die bloße Detektion von Rauschen hinausginge.

In einer vereinfachten Variante des erfindungsgemäßen Verfahrens kann die Anpassung des Spektrums dadurch deutlich vereinfacht werden, dass nur eine festgelegte Anzahl von Parametern der Vorlagefunktion verwendet werden, und die Anpassung im Hinblick auf diese Parameter erfolgt. Beispielsweise könnten Mittelwert und Streuung der Verteilung der Vorlagefunktion verwendet werden. Zur Anpassung werden gleichfalls Mittelwert und Streuung der Verteilung der Energieverteilungsfunktion ermittelt, und aus dem Vergleich dieser Parameter mit denen der Vorlagefunktion wird eine lineare Transformation für die Energiepegel des Spektrums bestimmt. Durch die Anwendung dieser linearen Transformation ergibt sich ein modifiziertes Spektrum, in welchem der störende Effekt des Hintergrundrauschens deutlich verringert ist. Sofern die Anwendung einer linearen Transformation nicht genügt, kann z.B. eine Transformation höherer Ordnung verwendet werden, die aus dem Vergleich einer entsprechenden Anzahl von Parametern der Energieverteilungsfunktion und der Vorlagefunktion, z.B. höherer Momente der Verteilungen, bestimmt wird.

Das erfindungsgemäße Verfahren eignet sich nicht nur für die Störungsverringerung für akustische Signale, wie z.B. Sprachsignale; vielmehr kann es ebenso für Muster anderer Art verwendet werden, welches sich durch eine über einem ein- oder mehrdimensionalen Feld aufgetragene Merkmalsgröße beschreiben läßt. Mögliche Einsatzgebiete sind demgemäß z.B. die Zeichenerkennung in geschriebenem Text od.dgl., Rekonstruktion und/oder Auswertung von Bildern usf.

PATENTANSPRÜCHE

1. Verfahren zur Unterdrückung von Störrauschen in einem Signalfeld (S2) enthaltend eine Vielzahl von Signalkomponenten, welche jeweils einen Wert eines Signalpegels annehmen und über einem Ordinatenbereich (T,F) auftragbar sind,
bei welchem aus dem Signalfeld (S2) eine Verteilungsfunktion (P2(E)) bestimmt wird, welche als Funktion des Signalpegels zu jedem seiner möglichen Signalpegel-Argumentwerte (E) angibt, wie groß der Anteil jener Signalkomponenten ist, deren Signalpegel niedriger als der Argumentwert (E) ist,

dadurch gekennzeichnet, dass

aufgrund eines Vergleiches der Verteilungsfunktion (P2(E)) mit einer vorbestimmten Referenzverteilungsfunktion (P0(E)) die Signalpegelwerte des Signalfeldes modifiziert werden, wobei die Abfolge der Signalkomponenten hinsichtlich ihrer Energiepegel ungeändert bleibt sowie Signalkomponenten, deren ursprüngliche Signalpegel gleich sind, gleiche modifizierte Signalpegel zugewiesen werden,
wobei als Referenzverteilungsfunktion (P0) eine aus einer Verteilungsfunktion, die für einen Satz von Referenzmustern bestimmt worden ist, gewonnene Funktion verwendet wird.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** für die Modifizierung der Signalpegelwerte ausgehend von einer Aufteilung des Wertebereichs der Signalpegel in eine Anzahl von Pegelbereichen für jeden Pegelbereich

- zu einem diesen Pegelbereich repräsentierenden, ersten Pegel (E0) unter Anwendung der Verteilungsfunktion (P2) und des Werts der Referenzverteilungsfunktion an dem ersten Pegel (P0(E0)) ein zweiter Pegel ausgewählt wird, für welchen der Wert der Verteilungsfunktion (P2(E)) dem genannten Wert der Referenzverteilungsfunktion (P0(E0)) möglichst nahe kommt, und
- jenen Signalkomponenten, deren Signalpegel zwischen dem ersten und dem zweiten Pegel fällt, der Wert des ersten Pegels (E0) zugewiesen wird.

3. Verfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** es für ein als zeit- und/oder frequenzabhängiges Spektrum eines akustischen Signals realisiertes Signalfeld ausgeführt wird.

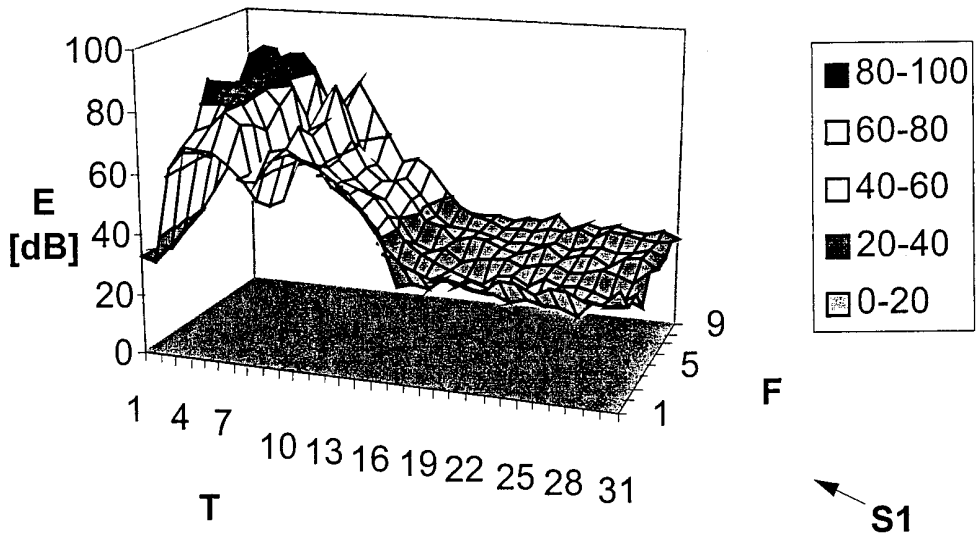

100
80
60
E
[dB]
40
20
0
1 4 7 10 13 16 19 22 25 28 31
T
9
5
1
F
S1
80-100
60-80
40-60
20-40
0-20

Fig. 1

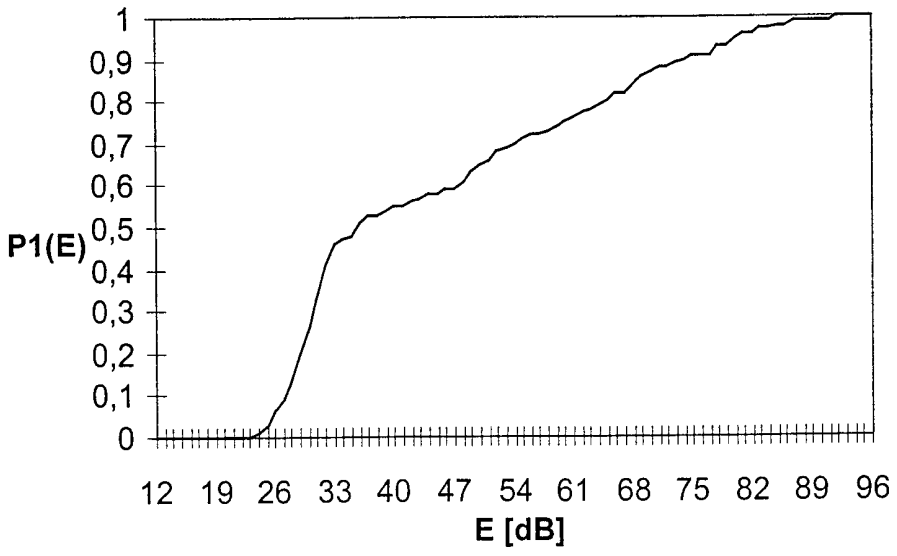
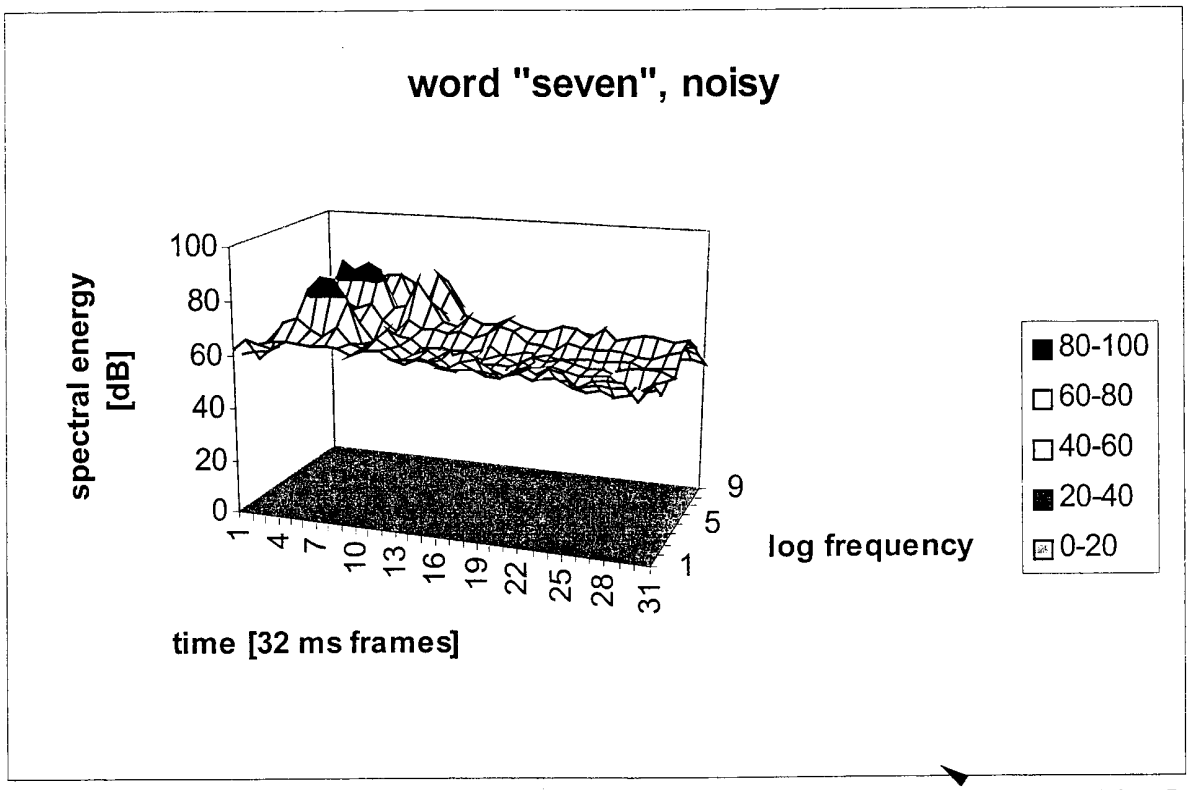

1
0,9
0,8
0,7
0,6
P1(E)
0,5
0,4
0,3
0,2
0,1
0
12 19 26 33 40 47 54 61 68 75 82 89 96
E [dB]

Fig. 2


word "seven", noisy
spectral energy [dB]
100
80
60
40
20
0
1
4
7
10
13
16
19
22
25
28
31
time [32 ms frames]
9
5
1
log frequency
80-100
60-80
40-60
20-40
0-20

S2 Fig. 3

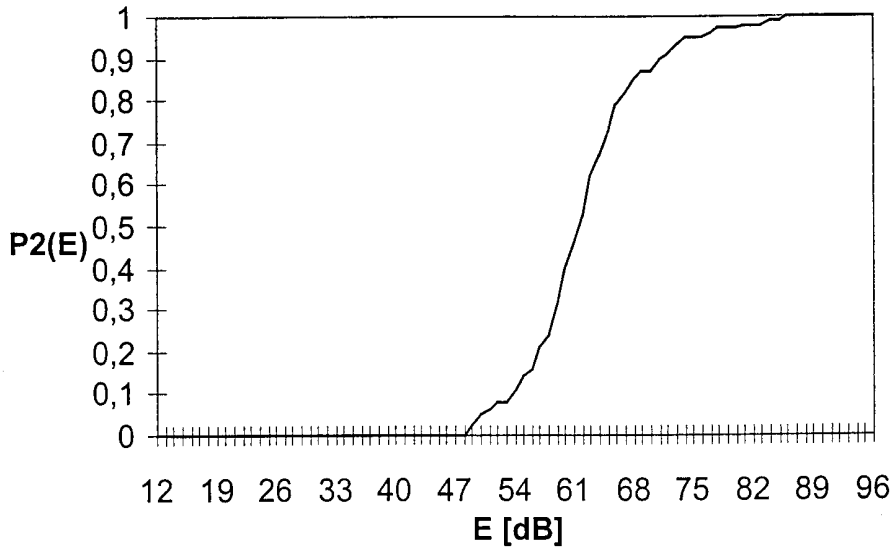
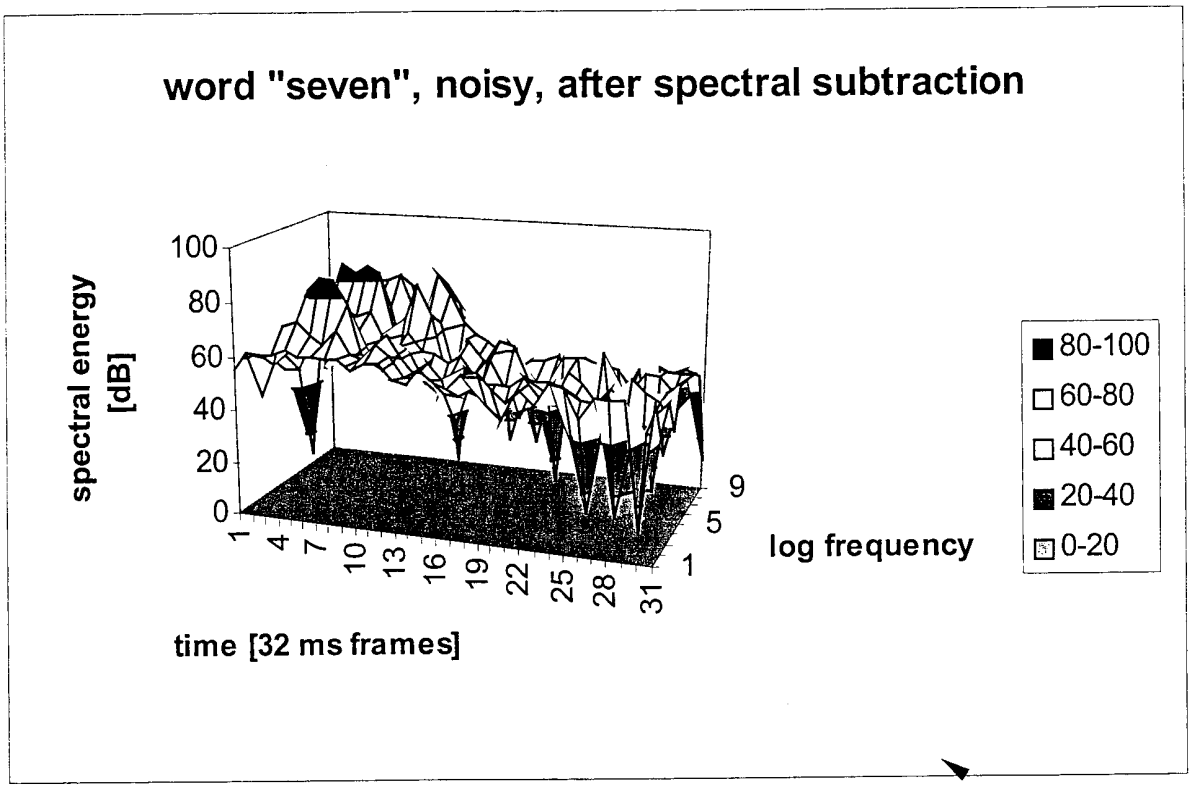

P2(E)
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
12 19 26 33 40 47 54 61 68 75 82 89 96
E [dB]

Fig. 4


word "seven", noisy, after spectral subtraction
spectral energy [dB]
100
80
60
40
20
0
1
4
7
10
13
16
19
22
25
28
31
time [32 ms frames]
1
5
9
log frequency
80-100
60-80
40-60
20-40
0-20

S3 Fig. 5

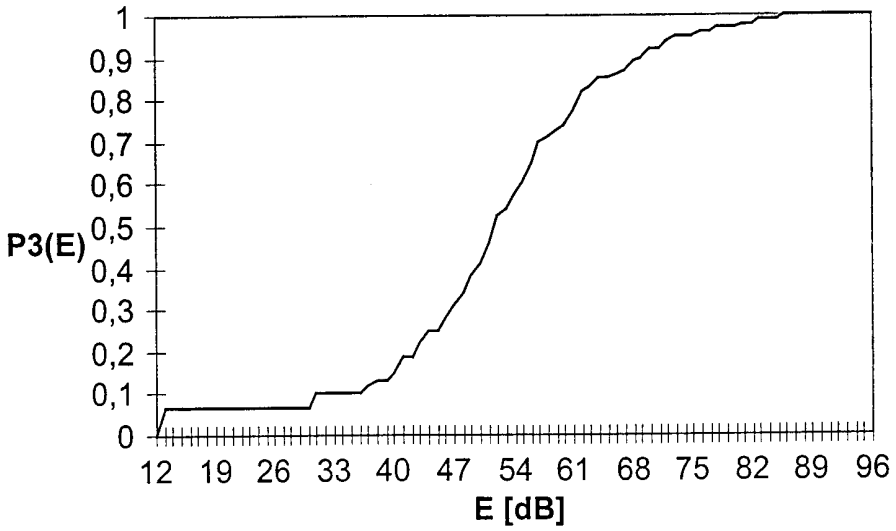

P3(E)
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
12 19 26 33 40 47 54 61 68 75 82 89 96
E [dB]

Fig. 6

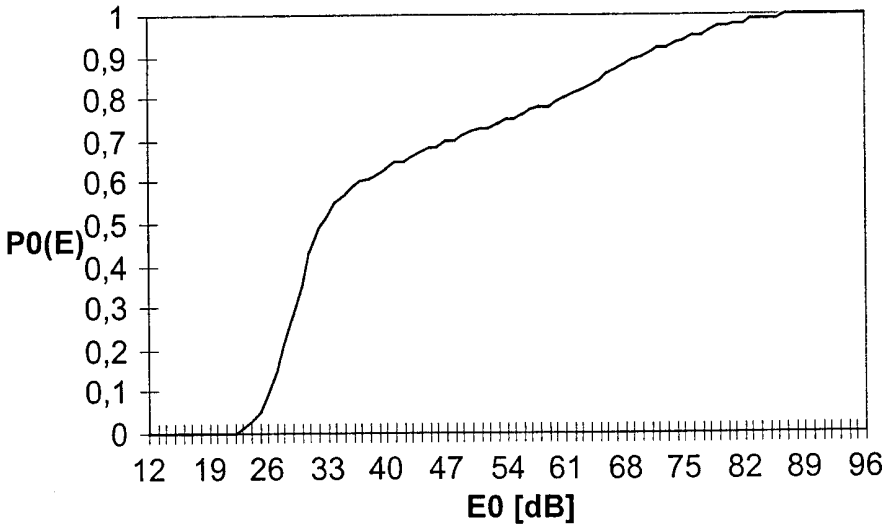

1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
P0(E)
12 19 26 33 40 47 54 61 68 75 82 89 96
E0 [dB]

Fig. 7

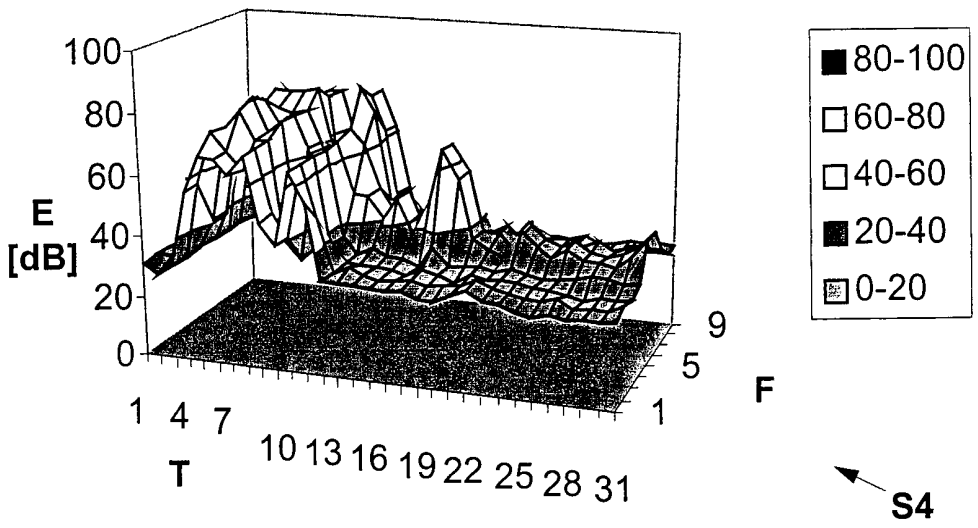

100
80
60
40
20
0
E
[dB]
1 4 7 10 13 16 19 22 25 28 31
T
1 5 9
F
80-100
60-80
40-60
20-40
0-20
S4

Fig. 8

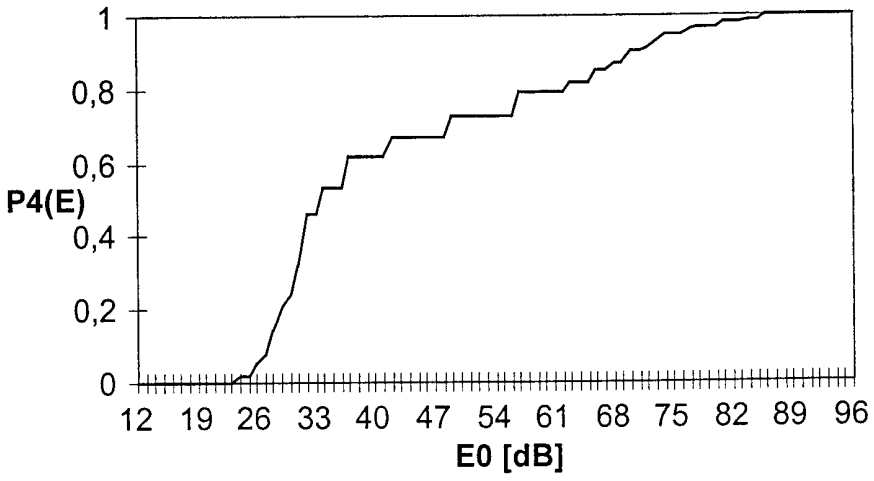

1
0,8
0,6
0,4
0,2
0
P4(E)
12 19 26 33 40 47 54 61 68 75 82 89 96
E0 [dB]

Fig. 9

INTERNATIONAL SEARCH REPORT

Int ational Application No
PCT/AT 00/00230

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L21/02 G06T5/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WHITE S A: "Restoration of nonlinearly distorted audio by histogram equalization" JOURNAL OF THE AUDIO ENGINEERING SOCIETY, NOV. 1982, USA, vol. 30, no. 11, pages 828-832, XP000978694 ISSN: 0004-7554 figure 2 --- -/--	1-3

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 February 2001	19/02/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Krembel, L

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	KENNY O P ET AL: "Time-frequency methods for enhancing speech" ADVANCED SIGNAL PROCESSING: ALGORITHMS, ARCHITECTURES, AND IMPLEMENTATIONS VII, SAN DIEGO, CA, USA, 28-30 JULY 1997, vol. 3162, pages 48-57, XP000979988 Proceedings of the SPIE - The International Society for Optical Engineering, 1997, SPIE-Int. Soc. Opt. Eng, USA ISSN: 0277-786X abstract paragraph '0003! ---	1-3
A	XUE-DONG YANG ET AL: "DIRECT MAPPING BETWEEN HISTOGRAMS: AN IMPROVED INTERACTIVE IMAGE ENHANCEMENT METHOD" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS,US,NEW YORK, IEEE, vol. _, 13 October 1991 (1991-10-13), pages 243-247, XP000238458 ISBN: 0-7803-0233-8 abstract paragraph 'OIII! ---	1
A	WHIPPLE G: "LOW RESIDUAL NOISE SPEECH ENHANCEMENT UTILIZING TIME-FREQUENCY FILTERING" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, ANDSIGNAL PROCESSING. (ICASSP),US,NEW YORK, IEEE, 19 April 1994 (1994-04-19), pages I-5-I-08, XP000529358 ISBN: 0-7803-1775-9 paragraph '0002! -----	1

1

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen
PCT/AT 00/00230

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 7 G10L21/02 G06T5/40

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
IPK 7 G10L G06T

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)
EPO-Internal, WPI Data, PAJ, INSPEC

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	WHITE S A: "Restoration of nonlinearly distorted audio by histogram equalization" JOURNAL OF THE AUDIO ENGINEERING SOCIETY, NOV. 1982, USA, Bd. 30, Nr. 11, Seiten 828-832, XP000978694 ISSN: 0004-7554 Abbildung 2 --- -/--	1-3

[X] Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

[ ] Siehe Anhang Patentfamilie

° Besondere Kategorien von angegebenen Veröffentlichungen :
"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts
1. Februar 2001	19/02/2001

Name und Postanschrift der Internationalen Recherchenbehörde	Bevollmächtigter Bediensteter
Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Krembel, L

1

Formblatt PCT/ISA/210 (Blatt 2) (Juli 1992)

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen
PCT/AT 00/00230

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	KENNY O P ET AL: "Time-frequency methods for enhancing speech" ADVANCED SIGNAL PROCESSING: ALGORITHMS, ARCHITECTURES, AND IMPLEMENTATIONS VII, SAN DIEGO, CA, USA, 28-30 JULY 1997, Bd. 3162, Seiten 48-57, XP000979988 Proceedings of the SPIE - The International Society for Optical Engineering, 1997, SPIE-Int. Soc. Opt. Eng, USA ISSN: 0277-786X Zusammenfassung Absatz '0003! ---	1-3
A	XUE-DONG YANG ET AL: "DIRECT MAPPING BETWEEN HISTOGRAMS: AN IMPROVED INTERACTIVE IMAGE ENHANCEMENT METHOD" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS,US,NEW YORK, IEEE, Bd. _, 13. Oktober 1991 (1991-10-13), Seiten 243-247, XP000238458 ISBN: 0-7803-0233-8 Zusammenfassung Absatz 'OIII! ---	1
A	WHIPPLE G: "LOW RESIDUAL NOISE SPEECH ENHANCEMENT UTILIZING TIME-FREQUENCY FILTERING" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, ANDSIGNAL PROCESSING. (ICASSP),US,NEW YORK, IEEE, 19. April 1994 (1994-04-19), Seiten I-5-I-08, XP000529358 ISBN: 0-7803-1775-9 Absatz '0002! -----	1

1