

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 1 212 751 B1

(12)

EUROPÄISCHE PATENTSCHRIFT

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:

27.10.2004 Patentblatt 2004/44

(21) Anmeldenummer: **00958032.5**

(22) Anmeldetag: **28.08.2000**

(51) Int Cl.7: **G10L 21/02, G06T 5/40**

(86) Internationale Anmeldenummer:

PCT/AT2000/000230

(87) Internationale Veröffentlichungsnummer:

WO 2001/020598 (22.03.2001 Gazette 2001/12)

(54) **VERFAHREN ZUR UNTERDRÜCKUNG VON STÖRRAUSCHEN IN EINEM SIGNALFELD**

METHOD FOR SUPPRESSING SPURIOUS NOISE IN A SIGNAL FIELD

PROCEDE D'ATTENUATION DE BRUITS PARASITES DANS UN CHAMP DE SIGNAL

(84) Benannte Vertragsstaaten:

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

(30) Priorität: **10.09.1999 AT 155999**

(43) Veröffentlichungstag der Anmeldung:

12.06.2002 Patentblatt 2002/24

(73) Patentinhaber: **SIEMENS**

**AKTIENGESELLSCHAFT
80333 München (DE)**

(72) Erfinder: **TSCHIRK, Wolfgang**

A-1150 Wien (AT)

(74) Vertreter: **Matschnig, Franz, Dipl.-Ing.**

**Patentanwalt,
Siebensterngasse 54,
Postfach 252
1071 Wien (AT)**

(56) Entgegenhaltungen:

- **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**
- **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 48-56, XP000979988 Proceedings of the SPIE - The International Society for Optical Engineering, 1997, SPIE-Int. Soc. Opt. Eng, USA ISSN: 0277-786X**
- **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**
- **WHIPPLE G: "LOW RESIDUAL NOISE SPEECH ENHANCEMENT UTILIZING TIME-FREQUENCY FILTERING" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. (ICASSP),US,NEW YORK, IEEE, 19. April 1994 (1994-04-19), Seiten I-5-I-08, XP000529358 ISBN: 0-7803-1775-9**

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).

EP 1 212 751 B1

Beschreibung

[0001] 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 Ordina-
 5 tenbereich auftragbar sind, bei welchem aus dem Signalfeld eine Verteilungsfunktion bestimmt wird, welche als Funk-
 tion des Signalpegels zu jedem seiner möglichen Signalpegel-Argumentwerte angibt, wie groß der Anteil jener Signal-
 komponenten ist, deren Signalpegel niedriger als der Argumentwert ist, die Signalpegelwerte des Signalfeldes so
 10 modifiziert werden, dass die Verteilungsfunktion des modifizierten Signalfeldes gleich einer vorbestimmten Referenz-
 verteilung ist, wobei die Abfolge der Signalkomponenten hinsichtlich ihrer Energiepegel ungeändert bleibt sowie Si-
 gnalkomponenten, deren ursprüngliche Signalpegel gleich sind, gleiche modifizierte Signalpegel zugewiesen werden,
 und als Referenzverteilungsfunktion eine aus einer Verteilungsfunktion, die für einen Satz von Referenzmustern be-
 stimmt worden ist, gewonnene Funktion verwendet wird.

[0002] Signalfelder, auf die das erfindungsgemäße Verfahren sich bezieht, werden beispielsweise in Mustererken-
 nungssystemen zur Beschreibung der zu erkennenden Muster verwendet. Der bei der Erkennung eines Musters ab-
 15 laufende Vorgang kann gewöhnlich grob in die folgenden Schritte aufgeteilt werden: Erfassung des Musters, Vorver-
 arbeitung und Klassifizierung.

[0003] 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
 20 Formats. Hierher gehört auch die Umwandlung eines Signal/Dataformats, 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 vor-
 verstärkt und in ein elektrisches Sprachsignal in analoger oder digitalisierter Form umgesetzt.

[0004] Das so erfasste Muster wird der Vorverarbeitung zugeführt, die eine Reduktion der zu verarbeitenden Daten
 25 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
 30 (Diskretisierung) und einer nachfolgenden, jeweils innerhalb eines Zeitrahmens durchgeführten Fourier-Transforma-
 tion 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.

[0005] 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 ge-
 35 ordnet, wobei diese Ordnung mit Hilfe eines oder mehrerer Ordinateparameter ausgedrückt ist. Beispielsweise be-
 steht ein als Zeit-Frequenz-Spektrum realisiertes Signalfeld aus vielen Spektralkomponenten, die jeweils einen eige-
 nen Energiepegel annehmen; die Spektralkomponenten sind nach Zeitrahmen und Frequenzband geordnet. Jeder
 Signalkomponente kann somit in dem Ordinatebereich, über den sich das Signalfeld erstreckt, ein eigener Bereich-
 selement des Ordinatebereichs zugeordnet werden, sodass die Bereichselemente insgesamt den Ordinatebereich
 40 des Signalfelds abdecken. In Abhängigkeit von der Anzahl der Ordinateparameter kann der Ordinatebereich ein-,
 zwei- oder mehrdimensional sein; dementsprechend sind die Bereichselemente Linien-, Flächen- oder (n -dimensio-
 nale) Volumenelemente.

[0006] Das durch die Vorverarbeitung erhaltene Signalfeld wird dem Klassifizierungssystem zugeführt. Dieses er-
 45 mittelt, zu welcher Erkennungsklasse - d.i. im Falle der Spracherkennung ein Wort eines vorgegebenen Wortschatzes
 oder eine Wortkette - eine Übereinstimmung gegeben ist. Das Erkennungsergebnis wird dann Ausgabe zugeführt, bei-
 spielsweise auf einer Anzeige, oder zur weiteren Verarbeitung genutzt, z.B. bei einer Befehlseingabe einer sprachori-
 entierten Einrichtung.

[0007] Die Ausführung einer Mustererkennung wird oftmals durch Störrauschen erschwert, das die zu erkennenden
 Muster überlagert. Beispielsweise kann die Leistungsfähigkeit eines Spracherkennungssystems durch akustischen
 50 Hintergrundlärm stark herabgesetzt oder ganz vereitelt werden.

[0008] 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
 55 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 Refe-
 renzrauschsignals 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.

[0009] 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 konvolute Überlagerung darstellen, wie z.B. Mischformen aus additiven und konvolutiven Störungen, sind noch schwieriger zu behandeln.

[0010] 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 Weibullverteilte Störung verlangen. Die Kenntnis der Verteilung oder zumindest der zugehörigen 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.

[0011] 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.

[0012] 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.

[0013] Aus keinen der genannten Dokumente geht die Verwendung einer aus Trainings- oder Referenzdaten gewonnenen Referenzverteilungsfunktion hervor.

[0014] Das Dokument WHITE SA: "Restoration of non linearly distorted audio by histogram equalization"n JOURNAL OF THE AUDIO ENGINEERING SOCIETY, Nov. 1982, USA, Bd 30, Nr. 11, Seiten 828-832, welches zur Bildung des Oberbegriffs des Anspruchs 1 herangezogen wurde, offenbart ein Verfahren zur Unterdrückung von Störgeräuschen, bei welchem die Amplitudenverteilung eines ungestörten Signals zu Hilfe genommen wird. Allerdings wird dabei eine komplexe Modellierung des logarithmischen Histogramms vorausgesetzt.

[0015] Es ist daher Aufgabe der Erfindung, ein Verfahren zur Rauschunterdrückung aufzuzeigen, das die Beeinträchtigung des Signalfelds durch das Störgeräuschen 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.

[0016] Die Aufgabe wird von einem Verfahren der eingangs genannten Art gelöst, bei welchem erfindungsgemäß 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 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 den ersten und den zweiten Pegel fällt, der Wert des ersten Pegels zugewiesen wird.

[0017] 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.

[0018] 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.

[0019] Es sei darauf hingewiesen, dass sich aus den oben genannten Schriften keinerlei Hinweise entnehmen las-

sen, dass eine Modifikation anhand einer Referenzverteilungsfunktion ohne Berücksichtigung der Art des Störrauschens erfolgreich sein könnte.

[0020] 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.

[0021] 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.

[0022] Des weiteren wird eine besonders zweckmäßige Realisierung der Erfindung für ein als zeitund/oder frequenzabhängiges Spektrum eines akustischen Signals realisiertes Signalfeld ausgeführt.

[0023] 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 zugehörnde Energieverteilungsfunktion einer Äußerung mit Geräuschhintergrund;

Fig. 5 und 6 ein Spektrogramm und die zugehörnde 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 zugehörnde Energieverteilungsfunktion, welche sich aus dem Spektrogramm der Fig. 3 mittels der erfindungsgemäßen Rauschreduktion anhand der Referenzverteilungsfunktion der Fig. 7 ergeben.

[0024] 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.

[0025] 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.

[0026] 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.

[0027] 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).

[0028] Fig. 2 zeigt die Energieverteilungsfunktion $P_1(E)$ zu dem in Fig. 1 dargestellten Spektrum S_1 . 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 P_1 bei 48 dB den Wert 0.6, denn 60 % der Energiepegel des Spektrums S_1 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 zugehörigen 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 angegebenen Pegel E , geteilt durch die Gesamtzahl der Komponenten aller dieser Spektren, an.

[0029] Fig. 3 zeigt das Spektrogramm S_2 zu einer Äußerung des Wortes von demselben Sprecher bei einer Wagen Geschwindigkeit von 113 km/h (70 mph). Wie aus dem Vergleich der Spektrogramme S_1 und S_2 (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 zugehörige Energieverteilungsfunktion $P_2(E)$ ist in Fig. 4 dargestellt.

[0030] Die Energieverteilungsfunktionen P_1 und P_2 (Fig. 2 bzw. 4) zeigen, dass die spektrale Verteilung des rauschfreien Signals S_1 deutlich verschieden von jener des geräuschbehafteten Signals S_2 ist, in dem die Hintergrundenergie um ungefähr 40 dB höher liegen als im Falle des rauschfreien Signals.

[0031] 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

$$S'(T,F) = E_0 = s_s(E, E_r) = (E^b - \alpha E_r^b)^{1/b},$$

wobei

$$\begin{aligned} E &= S(T,F) \text{ und} \\ E_r &= S_r(T,F) \end{aligned}$$

"subtrahiert" wird. Die Rauschreduktion nach der spektralen Subtraktion wurde im Rahmen der weiter unten beschriebenen Versuche der Anmelderin für das Spektrum S_2 durchgeführt. In Fig. 5 und 6 sind das Spektrum $S_3 = s_s(S_2, S_r)$, das sich bei der Anwendung der spektralen Subtraktion auf das Spektrogramm S_2 ergibt, und die zugehörige Energieverteilungsfunktion P_3 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 S_2 gewonnenes Referenzrauschen S_r . Wie aus Fig. 5 und 6 ersichtlich ist, ist das Hintergrundrauschen um ca. 10 dB niedriger als im unbehandelten Signal S_2 , 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.

[0032] Da das als Referenzrauschsignal S_r verwendete Signal nur statistisch mit dem Rauschen übereinstimmt, welches als Hintergrund des verrauschten Signals S_2 vorliegt, erzielt die spektrale Subtraktion eine Reduktion des Rauschpegels nur an einzelnen Komponenten des sich ergebenden Spektrums S_3 . 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 Niederpegel-Anteilen ca. ab Zeitrahmen 20 zu erkennen.

[0033] Gemäß der Erfindung erfolgt die Rauschunterdrückung für das vorliegende Sprachsignal S_2 unter Verwen-

dung 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.

[0034] 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.

[0035] 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 $E_0 = S'(T, F)$ aufweisen, d.h. für alle Spektralkomponenten gilt die Anpassungsbedingung

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

[0036] Des weiteren soll die Abfolge der Komponenten hinsichtlich ihrer Energiepegel nicht geändert werden, d.h.

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

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

[0037] Die Rauschunterdrückung kann als Konsequenz der Anpassungsbedingung (1) durch eine Anpassungsfunktion $R(E)$ vollständig beschrieben werden, die jedem ursprünglichen Pegel E einen modifizierten Pegel $E_0 = R(E)$ 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. $R(E_1) \leq R(E_2)$ wenn $E_1 < E_2$. Erfindungsgemäß erfolgt dies Anpassung des Spektrums derart, dass für die zugeordnete Energieverteilungsfunktion gilt $P_0(E_0) = P(E)$. 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.

[0038] 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.

[0039] Zunächst (ab der Marke PS/S) wird die zugehörige 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.

[0040] 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'.

[0041] 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 zugehörige Energieverteilungsfunktion P4 ist in Fig. 9 wiedergegeben.

[0042] Zur Verringerung des Aufwands bei der Durchführung des erfindungsgemäßen Verfahrens kann jeweils ein Pegelbereich des ursprünglichen Spektrums derart gemeinsam behandelt werden, dass den zugehörigen Spektralkomponenten ein einheitlicher modifizierter Pegel zugewiesen wird. Dieser modifizierte Pegel wird in Bezug auf einen

repräsentativen 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.

[0043] 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.

[0044] 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.

[0045] 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.

[0046] 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.

[0047] 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.

[0048] 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.

[0049] 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 usw.

Patentansprüche

1. Verfahren zur Unterdrückung von Störgeräuschen 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, die Signalpegelwerte des Signalfeldes (S2) so modifiziert werden, dass die Verteilungsfunktion P4 (E) des modifizierten Signalfeldes (S4)

gleich einer vorbestimmten Referenzverteilung ($P_0(E)$) ist, wobei die Abfolge der Signalkomponenten hinsichtlich ihrer Energiepegel ungeändert bleibt sowie Signalkomponenten, deren ursprüngliche Signalpegel gleich sind, gleiche modifizierte Signalpegel zugewiesen werden, und als Referenzverteilungsfunktion (P_0) eine aus einer Verteilungsfunktion, die für einen Satz von Referenzmustern bestimmt worden ist, gewonnene Funktion verwendet wird,

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 (E_0) unter Anwendung der Verteilungsfunktion (P_2) und des Werts der Referenzverteilungsfunktion an dem ersten Pegel ($P_0(E_0)$) ein zweiter Pegel ausgewählt wird, für welchen der Wert der Verteilungsfunktion ($P_2(E)$) dem genannten Wert der Referenzverteilungsfunktion ($P_0(E_0)$) möglichst nahe kommt, und
- jenen Signalkomponenten, deren Signalpegel zwischen den ersten und den zweiten Pegel fällt, der Wert des ersten Pegels (E_0) zugewiesen wird.

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

Claims

1. A method for suppressing spurious noise in a signal field (S_2) containing a plurality of signal components which each adopt a value of a signal level and may be assigned to an ordinate range, where from the signal field (S_2) a distribution function ($P_2(E)$) is determined, which, as a function of the signal level, indicates for each of its possible argument values of signal level (E), the size of the fraction of those signal components whose signal level is lower than the argument value

the signal level values of the signal field (S_2) are modified in such a manner that the distribution function ($P_4(E)$) of the modified signal field (S_4) equals a predetermined reference distribution ($P_0(E)$), wherein the sequence of the signal components remain unchanged with regard to their energy level and signal components, whose original signal levels are identical, are assigned the same modified signal levels, and a function obtained from a distribution function that was determined for a set of reference models is used as reference distribution function (P_0),

characterized in that

to modify the signal level values, starting from a division of the value range of the signal levels into a number of level ranges, for each level range

- to a first level (E_0) representing the respective level range, using the distribution function (P_2) and the value of the reference distribution function at the first level ($P_0(E_0)$), a second level is selected for which the value of the distribution function ($P_2(E)$) is as close as possible to the said value of the reference distribution function ($P_0(E_0)$), and
- those signal components, whose signal levels fall between the first and the second level, are assigned the value of the first level (E_0).

2. The method of claim 1, **characterized in that** it is carried out for a signal field realized as a time and/or frequency dependent spectrum of an acoustic signal.

Revendications

1. Procédé pour supprimer des bruits parasites dans un champ de signal (S_2) contenant une multitude de composantes de signal qui adoptent chacune une valeur d'un niveau de signal et peuvent être portées sur une plage d'ordonnée (T, F), avec lequel est déterminée à partir du champ de signal (S_2) une fonction de distribution ($P_2(E)$) qui indique en fonction du niveau de signal à chacune de ses valeurs d'argument de niveau de signal (E) possibles, quelle est la proportion des composantes de signal dont le niveau de signal est inférieur à la valeur d'argument (E), et les valeurs de niveau de signal du champ de signal (S_2) sont modifiées de sorte que la fonction de distribution ($P_4(E)$) du champ de signal modifié (S_4) est égale à une distribution de référence ($P_0(E)$) prédéfinie, avec lequel l'ordre des composantes de signal reste inchangé du point de vue de leur niveau d'énergie, les composantes de signal dont les niveaux de signal originaux sont égaux sont associées à des niveaux de signal modifiés égaux, et comme fonction de distribution de référence (P_0) est utilisée une fonction générée à partir d'une fonction de dis-

tribution qui était destinée à un jeu d'échantillons de référence

caractérisé en ce que

Pour la modification des valeurs de niveaux de signal à partir d'un découpage de la plage de valeurs des niveaux de signal en un certain nombre de plages de niveau pour chaque plage de niveau

- à un premier niveau (E0) représentant cette plage de niveau, en utilisant la fonction de distribution (P2) et la valeur de la fonction de distribution de référence au premier niveau (P0(E0)), est choisi un deuxième niveau pour lequel la valeur de la fonction de distribution (P2(E)) se rapproche autant que possible de la valeur mentionnée de la fonction de distribution de référence (P0(E0)), et
- la valeur du premier niveau (E0) est attribuée aux composantes de signal dont le niveau de signal se trouve entre le premier et le deuxième niveau.

2. Procédé selon la revendication 1, **caractérisé en ce qu'il** est réalisé pour un champ de signal sous forme de spectre d'un signal acoustique variant en fonction du temps et/ou de la fréquence.

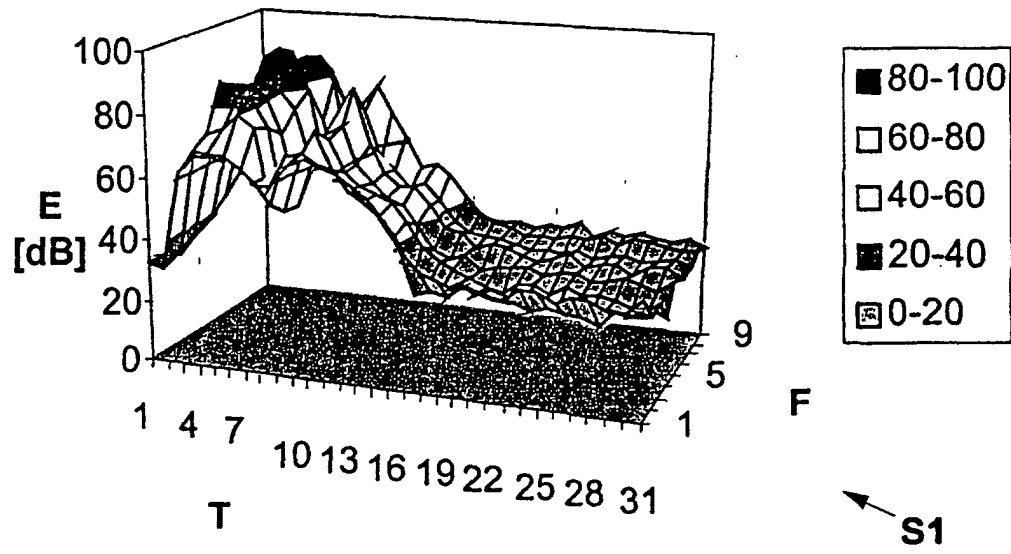


Fig. 1

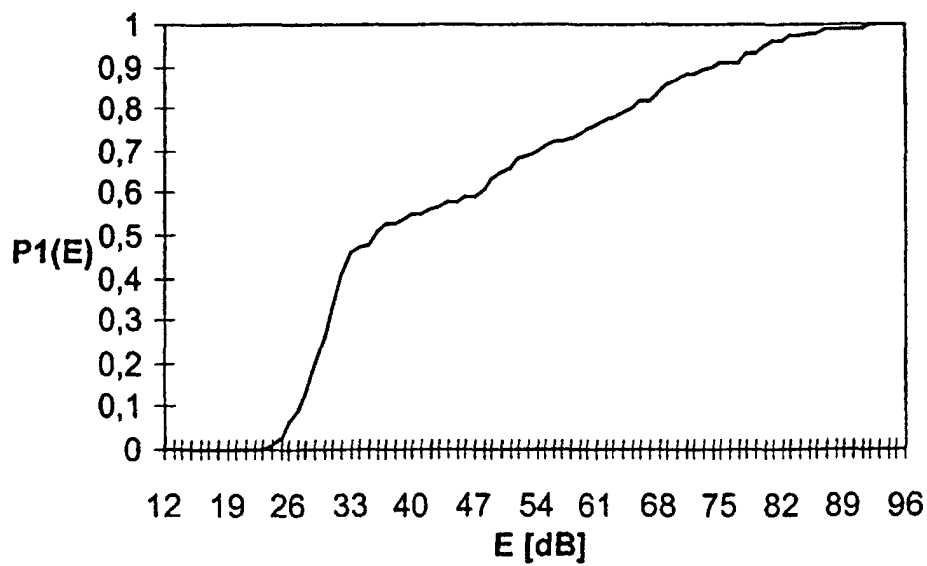
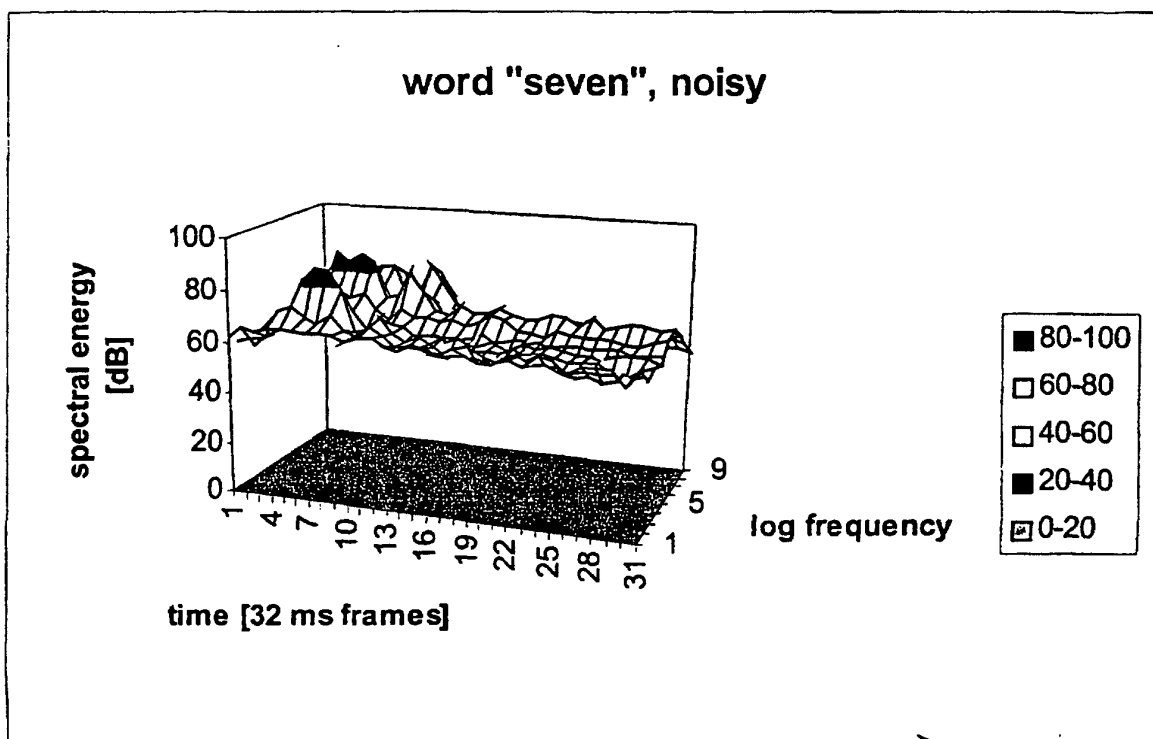


Fig. 2



S2 Fig. 3

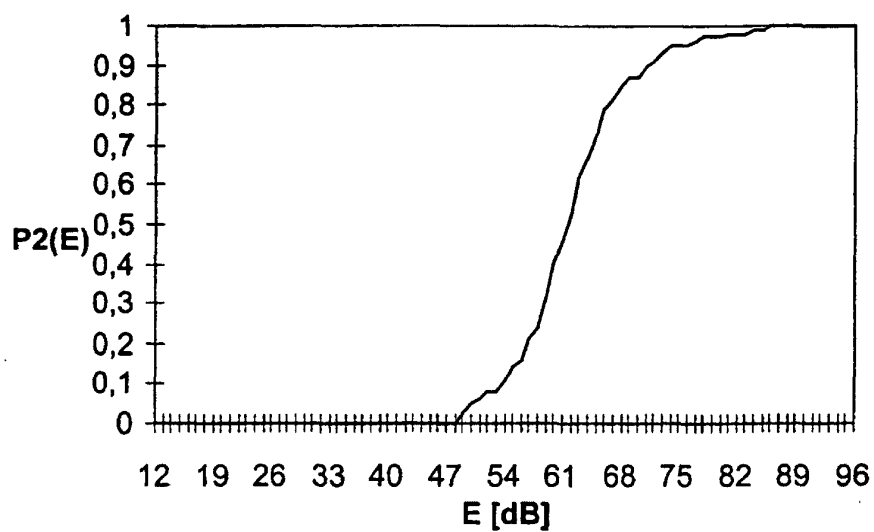
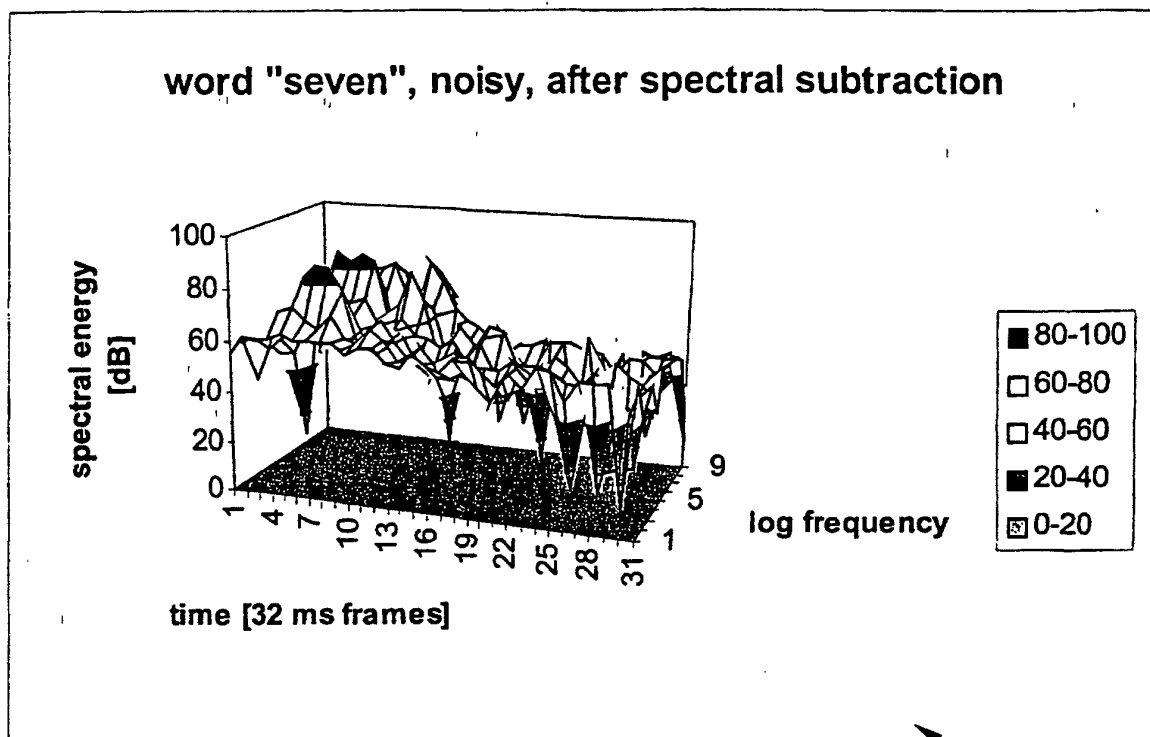


Fig. 4



S3 Fig. 5

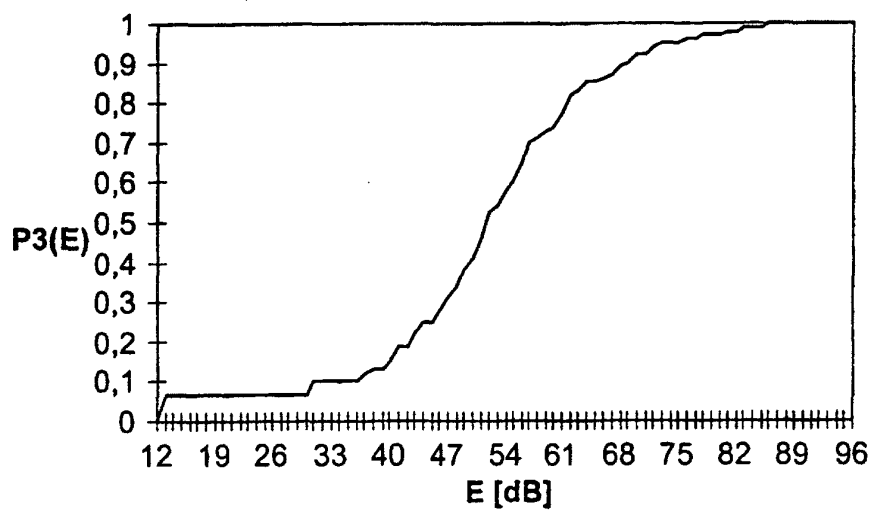


Fig. 6

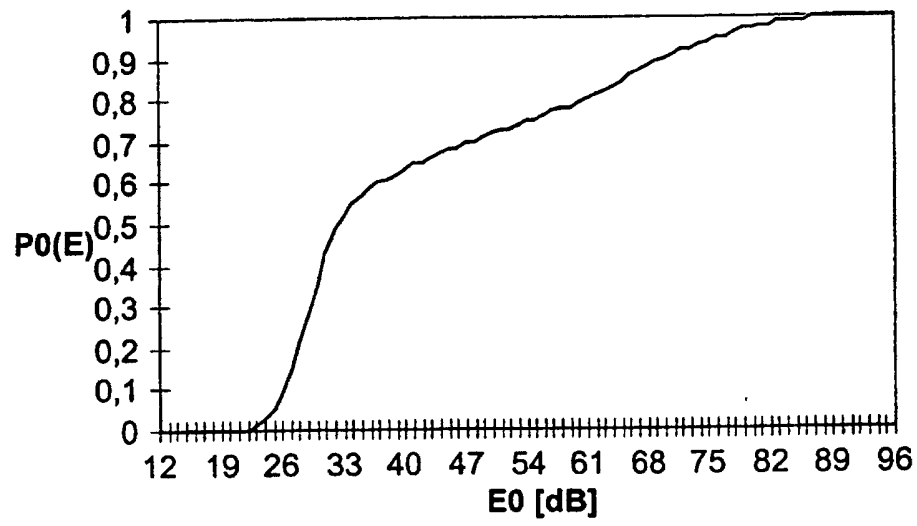


Fig. 7

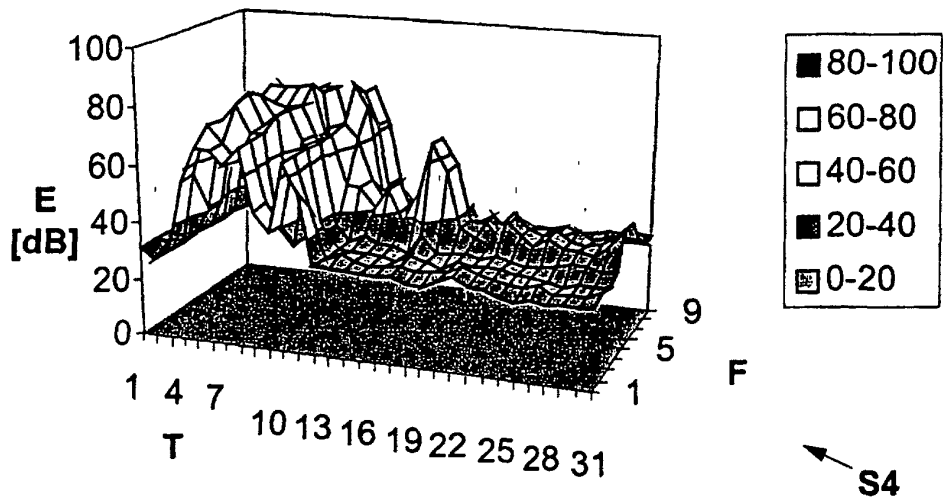


Fig. 8

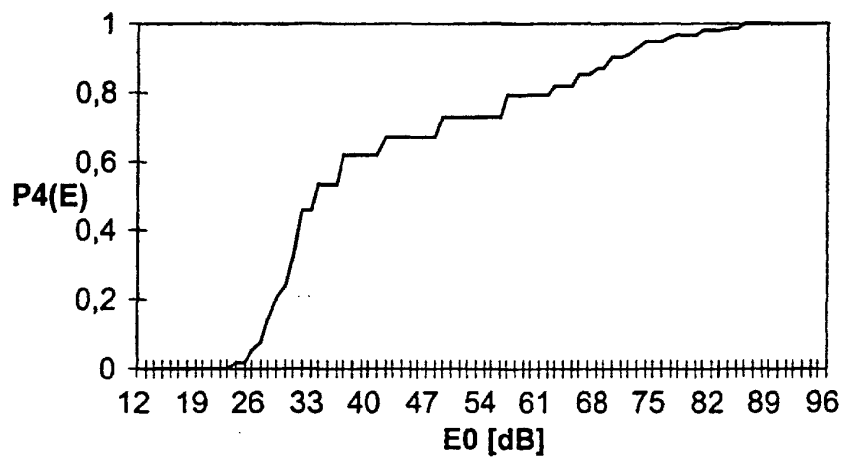


Fig. 9