

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 689 191 B1

(12)

EUROPÄISCHE PATENTSCHRIFT

(45) Veröffentlichungstag und Bekanntmachung des
Hinweises auf die Patenterteilung:
23.05.2001 Patentblatt 2001/21

(51) Int Cl.7: **G10L 21/02**

(21) Anmeldenummer: **95201578.2**

(22) Anmeldetag: **14.06.1995**

(54) Sprachverarbeitungsvorrichtung und Mobilfunkendgerät

Speech processing apparatus and mobile transceiver

Dispositif pour le traitement de paroles et émetteur-récepteur mobile

(84) Benannte Vertragsstaaten:
DE FR GB

(30) Priorität: **22.06.1994 DE 4421853**

(43) Veröffentlichungstag der Anmeldung:
27.12.1995 Patentblatt 1995/52

(73) Patentinhaber:
• **Philips Corporate Intellectual Property GmbH**
52064 Aachen (DE)
Benannte Vertragsstaaten:
DE
• **Koninklijke Philips Electronics N.V.**
5621 BA Eindhoven (NL)
Benannte Vertragsstaaten:
FR GB

(72) Erfinder: **Rainer, Martin, Dipl.-Ing., c/o Philips**
D-20097 Hamburg (DE)

(74) Vertreter: **Volmer, Georg, Dipl.-Ing. et al**
Philips Corporate Intellectual Property GmbH,
Habsburgerallee 11
52064 Aachen (DE)

(56) Entgegenhaltungen:
EP-A- 0 073 869 WO-A-91/03790
US-A- 3 997 772

- **IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, Bd. 29, Nr. 3, Juni 1981, US, Seiten 582-587, XP002028506 ETTER ET AL.: "Adaptive estimation of time delays in sampled data systems"**
- **INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 1979, 2. - 4. April 1979, WASHINGTON DC, US, Seiten 128-131, XP002028507 CHAN ET AL.: "A parameter estimation approach to time delay estimation"**
- **IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, Bd. 33, Nr. 4, August 1985, US, Seiten 812-822, XP002028508 SMITH ET AL.: "Adaptive multipath delay estimation"**

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 0 689 191 B1

Beschreibung

[0001] Die Erfindung betrifft ein Mobilfunkendgerät mit einer Sprachverarbeitungsvorrichtung.

[0002] Auf dem Gebiet der Sprachverarbeitung sind häufig in zu verarbeitenden Sprachsignalen Rauschsignalanteile enthalten, was zur Verringerung der Sprachqualität und damit insbesondere zu einer verschlechterten Sprachverständlichkeit führt. Dieses Problem tritt beispielsweise bei Mobilfunkendgeräten auf, die in Kraftfahrzeugen verwendet werden und eine Freisprecheinrichtung aufweisen. Sprachsignale, die von im Kraftfahrzeug angeordneten Mikrofonen der Freisprecheinrichtung empfangen werden, enthalten einerseits Sprachsignalanteile, die vom jeweiligen Benutzer (Sprachquelle) des Mobilfunkendgerätes innerhalb des Kraftfahrzeuges erzeugt werden, und andererseits Rauschsignalanteile, die aus sonstigen Umgebungsgeräuschen und während einer Fahrt im wesentlichen aus Motor- und Fahrgeräuschen bestehen.

[0003] Aus "IEEE Transactions on Acoustics, Speech, and Signal Processing, VOL. ASSP-29, No. 3, June 1981, pp. 582-587" ist eine Anordnung zur adaptiven Schätzung von Zeitverzögerungen von zwei stark korrelierten Signalen in digitalen Systemen beschrieben. Eines der beiden Signale wird von einem steuerbaren Verzögerungsglied verzögert. Die Verzögerungswerte des Verzögerungsgliedes werden adaptiv an die korrelierten Signale angepaßt. Die Bestimmung der Verzögerungswerte erfolgt mit Hilfe eines Algorithmus, der mittlerweile von der Fachwelt als LMS-Algorithmus (Least Mean Square) bezeichnet wird. Dieser Algorithmus beruht auf der Minimierung der Leistung bzw. des Quadrates von Fehlerwerten, die sich durch Bildung der Differenz von dem verzögerten und dem nicht verzögerten Signal ergeben. Kern des LMS-Algorithmus ist die rekursive Berechnung der Verzögerungswerte mit Hilfe von Schätzwerten für den Gradienten der Leistung der Fehlerwerte.

[0004] Im oben zitierten Stand der Technik wird zur Bildung der Fehlerwerte jeweils die Differenz zweier Abtastwerte von zwei gegeneinander zeitversetzten Signalen gebildet, wobei eines der Signale verzögert wird. Der entsprechende Verzögerungswert ist auf ein ganzzahliges Vielfaches eines Abtastintervalls der Signale gerundet. Dabei treten Konvergenzprobleme derart auf, daß beim Erreichen sehr kleiner Fehlerwerte starke Oszillationen der gerundeten Verzögerungswerte auftreten. Die Verzögerungswerte oszillieren dabei zwischen zwei gerundeten Verzögerungswerten im Abstand eines Abtastintervalls.

[0005] Der Erfindung liegt die Aufgabe zugrunde, die Sprachqualität der zu verarbeitenden Sprachsignale zu verbessern und Konvergenzprobleme zu verringern.

[0006] Die Aufgabe wird dadurch gelöst, daß die Sprachverarbeitungsvorrichtung zur Verarbeitung eines ersten und mindestens eines weiteren aus Rausch- und Sprachsignalanteilen bestehenden und als Abtastwerte vorliegenden Sprachsignals vorgesehen ist, daß Verzögerungsmittel zur Verzögerung des abgetasteten weiteren Sprachsignals vorgesehen sind, daß Steuermittel

- zur Bildung von Gradientenschätzwerten durch Multiplikation von Fehlerwerten für zwei Sprachsignale mit den Ausgangswerten eines Digitalfilters, das eine Phasenverschiebung von 90 Grad bewirkt und zur Filterung eines der zwei Sprachsignale dient,
- zur rekursiven Ermittlung von Verzögerungsschätzwerten aus den Gradientenschätzwerten, wobei aus den Verzögerungsschätzwerten durch Rundung die Verzögerungswerte gebildet werden, die zur Einstellung der Verzögerungsmittel dienen und
- zur Bildung jeweils wenigstens eines Fehlerwertes für einen bestimmten Abtastzeitpunkt aus der Differenz zwischen einem Sprachsignalschätzwert, der zur Abschätzung des weiteren Sprachsignals zu einem gegenüber dem bestimmten Abtastzeitpunkt um den Verzögerungsschätzwert verschobenen Zeitpunkt dient und durch Interpolation von Abtastwerten des weiteren Sprachsignals gebildet wird, und dem Abtastwert eines anderen der zu verarbeitenden Sprachsignale zu dem bestimmten Abtastzeitpunkt

vorgesehen sind und daß eine Addiervorrichtung zum Addieren der gegeneinander zeitversetzten Sprachsignale vorgesehen ist.

[0007] Die Gradientenschätzwerte dienen zur Abschätzung des jeweiligen Gradienten der Leistung der Fehlerwerte oder anders ausgedrückt der quadrierten Fehlerwerte. Die Steuermittel bestimmen die Verzögerungsschätzwerte derart, daß die Leistung der Fehlerwerte verringert wird. Dabei wird die Konvergenz der aus den Verzögerungsschätzwerten ermittelten Verzögerungswerte erheblich verbessert, da die Verzögerungsschätzwerte gegenüber den Verzögerungswerten aufgrund der Rundung eine höhere Auflösung aufweisen. Oszillationen der Verzögerungswerte werden so im wesentlichen vermieden. Die Auflösung der Verzögerungswerte ist gegenüber der Auflösung der Verzögerungsschätzwerte geringer gewählt, um den technischen Aufwand beim Verzögern der Sprachsignale möglichst gering zu halten. Das Signal-/ Rauschleistungsverhältnis und die Sprachqualität eines am Ausgang der Addiervorrichtung anliegenden Summensignals sind gegenüber dem Signal-/ Rauschleistungsverhältnis und der Sprachqualität der einzelnen Sprachsignale verbessert.

[0008] In einer Ausgestaltung der Erfindung ist das Digitalfilter ein digitaler Hilbert-Transformator.

[0009] Ein digitaler Hilbert-Transformator, der eine Phasenverschiebung von 90 Grad für alle Frequenzen bewirkt, besitzt betragsmäßig die Übertragungsfunktion eines Tiefpasses, so daß insbesondere für die tiefen und für ein Sprachsignal wesentlichen Frequenzen die gerundeten Verzögerungswerte gut konvergieren. Der Hilbert-Transformator kann beispielsweise auch durch einen Differenzierer ersetzt werden, der ebenfalls eine Phasenverschiebung von 90 Grad bewirkt. Allerdings hat ein Differenzierer betragsmäßig eine linear ansteigende Übertragungsfunktion, so daß insbesondere die tiefen Frequenzen eines Sprachsignals unterdrückt werden, so daß sich keine so gute Konvergenz wie bei einem Hilbert-Transformator ergibt.

[0010] In einer anderen Ausgestaltung sind Mittel zur Glättung der Gradientenschätzwerte vorgesehen.

[0011] Damit ergibt sich eine verbesserte Schätzung der Verzögerungsschätzwerte.

[0012] In einer weiteren Ausgestaltung ist die Sprachverarbeitungsvorrichtung zur Verarbeitung von drei Sprachsignalen vorgesehen.

[0013] Gegenüber einer Sprachverarbeitungsvorrichtung zur Verarbeitung von nur zwei Sprachsignalen läßt sich auf diese Weise das Signal-/ Rauschleistungsverhältnis und die Sprachqualität des am Ausgang der Addiervorrichtung anliegenden Summensignals verbessern.

[0014] Die Erfindung kann weiterhin dadurch ausgestaltet werden, daß zur Ermittlung eines Verzögerungsschätzwertes für das weitere Sprachsignal die Verwendung einer Linearkombination von Fehlerwerten vorgesehen ist.

[0015] Auf diese Weise wird die Stabilität der Sprachverarbeitungsvorrichtung erhöht.

[0016] Für eine andere Ausgestaltung der Erfindung sind Verzögerungsmittel zur Verzögerung des ersten Sprachsignals mit einer festen Verzögerungszeit vorgesehen.

[0017] Ohne die eine feste Verzögerung bewirkenden Verzögerungsmittel sind nur Zeitversätze zwischen dem ersten und dem/den weiteren Sprachsignal(en) einstellbar, mit denen ein Vorlaufen des ersten Sprachsignals bewirkt wird. Je nach Position einer die Sprachsignalanteile erzeugenden Sprachquelle gegenüber Mikrofonen der Sprachverarbeitungsvorrichtung, die zur Umwandlung der von der Sprachquelle erzeugten akustischen Sprachsignale in elektrische Sprachsignale dienen, muß allerdings auch ein Nacheilen des ersten Sprachsignal einstellbar sein, was mit Hilfe dieser Ausgestaltung auf einfache Weise realisierbar ist.

[0018] Zur weiteren Ausgestaltung der Erfindung ist die Sprachverarbeitungsvorrichtung in eine Freisprecheinrichtung integriert.

[0019] Insbesondere bei Freisprecheinrichtungen besteht das Problem, daß empfangene Sprachsignale störende Rauschsignalanteile aufweisen, die das Signal-/ Rauschleistungsverhältnis und die Sprachqualität der Sprachsignale verschlechtern. Gerade bei Mobilfunkendgeräten tritt dieses Problem auf, wenn diese in einer stark verrauschten Umgebung eingesetzt werden, wie z.B. in einem Automobil. Die Verwendung der beschriebenen Erfindung bewirkt deshalb gerade beim Einsatz in Freisprecheinrichtungen eine verbesserte Kommunikation zwischen den Gesprächsteilnehmern.

[0020] Ausführungsbeispiele werden nachstehend anhand der Zeichnungen näher erläutert.

[0021] Es zeigen:

- Fig. 1 eine Sprachverarbeitungsvorrichtung für zwei Sprachsignale,
- Fig. 2 eine Steuervorrichtung zur Einstellung eines Zeitversatzes zwischen den beiden Sprachsignalen nach Fig. 1,
- Fig. 3 eine Sprachverarbeitungsvorrichtung für drei Sprachsignale,
- Fig. 4 und 5 Blockschaltbilder mit Steuervorrichtungen zur Einstellung von Zeitversätzen zwischen den drei Sprachsignalen nach Fig. 3,
- Fig. 6 und 7 ein Blockschaltbild und ein Flußdiagramm zur Bestimmung des Signal-/ Rauschleistungsverhältnisses eines Sprachsignals,
- Fig. 8 eine Einteilung von geglätteten Leistungswerten eines Sprachsignals in Gruppen und Untergruppen und
- Fig. 9 ein Mobilfunkendgerät mit einer Sprachverarbeitungsvorrichtung nach Fig. 1 bis 8.

[0022] Die in Fig. 1 dargestellte Sprachverarbeitungsvorrichtung enthält zwei Mikrophone M1 und M2. Diese dienen zur Umwandlung von akustischen in elektrische Sprachsignale, die sich aus Sprach- und Rauschsignalanteilen zusammensetzen. Die Sprachsignalanteile stammen von einer einzelnen Sprachquelle (Sprecher), die im Regelfall unterschiedliche Abstände zu den beiden Mikrofonen M1 und M2 aufweist. Die Sprachsignalanteile sind somit in hohem Maße korreliert. Die Rauschsignalanteile der beiden von den Mikrofonen M1 und M2 empfangenen Sprachsignale sind nicht von der einzelnen Sprachquelle erzeugte Umgebungsgeräusche, die bei geeigneten Mikrofonabständen im Bereich von 10 bis 60 cm als unkorreliert oder nur wenig korreliert vorausgesetzt werden können, wenn sich die Mikrophone in einer sogenannten verhallten Umgebung wie beispielsweise im Auto oder in einem Büro befinden. Befinden sich Sprachquelle und Sprachverarbeitungsvorrichtung beispielsweise in einem Kraftfahrzeug, werden die Rauschsignalanteile insbesondere durch Motor- und Fahrgeräusche verursacht.

[0023] Die von den Mikrofonen M1 und M2 erzeugten Mikrophonsignale werden von Analog-Digitalumsetzern 1 und 2 digitalisiert. Die sich ergebenden digitalisierten und damit als Abtastwerte $x1(i)$ und $x2(i)$ vorliegenden Mikrophonsignale werden von einer Steuervorrichtung 3 ausgewertet, die zur Steuerung und Einstellung eines Verzögerungsgliedes 4 dient. Die abgetasteten Mikrophonsignale $x1(i)$ und $x2(i)$ werden im folgenden abgekürzt als Mikrofon- oder Sprachsignale bezeichnet. Das Verzögerungsglied 4 verzögert das Mikrophonsignal $x1$ mit durch die Steuervorrichtung 3 einstellbaren Verzögerungswerten $T1$. Eine Addiervorrichtung 5 addiert das vom Verzögerungsglied 4 verzögerte Mikrophonsignal $x1(i)$ und das von einem Verzögerungsglied 16 mit einer konstanten Zeitverzögerung $T_{\max}$ verzögerte Mikrophonsignal $x2(i)$. Das Verzögerungsglied 16 ist vorgesehen, um sowohl ein Vorlaufen als auch ein Nacheilen des Mikrophonsignals $x1(i)$ gegenüber dem Mikrophonsignal $x2(i)$ einstellen zu können. Ein am Ausgang der Addiervorrichtung 5 anliegendes Summensignal $X(i)$ ist ein abgetastetes Sprachsignal, dessen Signal-/ Rauschleistungsverhältnis gegenüber den Signal-/ Rauschleistungsverhältnissen der Sprachsignale $x1(i)$ und $x2(i)$ erhöht ist. Durch eine geeignete Einstellung der Verzögerungszeit $T1$ des Verzögerungsgliedes 4 wird bei der Addition durch die Addiervorrichtung 5 eine Verstärkung der Leistung der Sprachsignalanteile der beiden Sprachsignale $x1(i)$ und $x2(i)$ ungefähr um den Faktor 4 und eine Verstärkung der Leistung der Rauschsignalanteile nur ungefähr um den Faktor 2 bewirkt. Damit ergibt sich eine Verbesserung des leistungsbezogenen Signal-/ Rauschleistungsverhältnisses von ungefähr 3 dB.

[0024] In Fig. 2 wird die Funktionsweise der Steuervorrichtung 3 anhand eines Blockschaltbildes näher erläutert. Aus dem Sprachsignal $x2(i)$ und Sprachsignalschätzwerten $x1_{\text{int}}(i)$ ergeben sich Fehlerwerte $e_{12}(i)$ durch Differenzbildung nach

$$e_{12}(i) = x1_{\text{int}}(i) - x2(i) \quad (1)$$

[0025] Die Sprachsignalschätzwerte $x1_{\text{int}}(i)$ sind Werte, die sich aus einer Interpolation von Abtastwerten des Sprachsignals $x1(i)$ ergeben. Die Bestimmung der Sprachsignalschätzwerte $x1_{\text{int}}(i)$ wird später erläutert. i ist eine Variable, die ganzzahlige Werte annehmen kann und mit der einerseits Abtastzeitpunkte der Sprachsignale $x1(i)$ und $x2(i)$ und andererseits auch Programmzyklen der programmierbaren und Steuermittel aufweisenden Steuervorrichtung 3 indiziert werden, wobei in einem Programmzyklus jeweils ein neuer Abtastwert per Sprachsignal verarbeitet wird.

[0026] Ein digitales Filter 6 führt eine Hilbert-Transformation der Abtastwerte $x2(i)$ durch:

$$x2_H(i) = \sum_{k=0}^K h(k) * x2(i-k) \quad (2)$$

[0027] Das die Werte $x2_H(i)$ von $x2(i)$ liefernde Digitalfilter 6 ist ein FIR-Filter der Ordnung K , das Koeffizienten $h(0)$, $h(1)$, ..., $h(K)$ aufweist. Im vorliegenden Ausführungsbeispiel ist K gleich sechzehn, so daß das Digitalfilter 6 siebzehn Koeffizienten aufweist. Das Digitalfilter 6 besitzt dem Betrage nach die Übertragungsfunktion eines Tiefpasses. Es erzeugt weiterhin eine Phasenverschiebung von 90 Grad. Die feste Phasenverschiebung von 90 Grad ist die entscheidende Eigenschaft des Digitalfilters 6, der Verlauf des Betrages der Übertragungsfunktion ist für das Funktionieren der Sprachverarbeitungsvorrichtung nicht entscheidend. So kann das Digitalfilter 6 auch mit Hilfe eines Differenzierers realisiert werden, was allerdings zu einer Unterdrückung von niederfrequenten Anteilen von $x2(i)$ und damit zu einer verringerten Leistungsfähigkeit der Sprachverarbeitungsvorrichtung führen würde.

[0028] Die Ausgangswerte $x2_H(i)$ werden mit den Fehlerwerten $e_{12}(i)$ und dem Kehrwert $1/P_{x2}(i)$ einer Kurzzeitleistung $P_{x2}(i)$ multipliziert, wobei die Kurzzeitleistung $P_{x2}(i)$ nach

$$P_{x2}(i) = P_{x2}(i-1) + [x2(i)]^2 - [x2(i-N)]^2 \quad (3)$$

gebildet wird. N gibt die Anzahl der in die Berechnung eingehenden Abtastwerte von $x2$ an. N ist beispielsweise gleich 65. Die Multiplikation mit $1/P_{x2}(i)$ dient zur Vermeidung von Instabilitäten in der Steuervorrichtung 3 beim Steuern des Verzögerungsgliedes 4. Damit ergibt sich durch

$$\text{grad}(i) = \frac{1}{P_{x2}(i)} * e(i) * x2_H(i) \quad (4)$$

ein auf die Kurzzeitleistung $P_{x2}(i)$ normierter geschätzter Gradient $\text{grad}(i)$ der Quadrate bzw. der Leistung der Fehlerwerte $e_{12}(i)$ im Programmzyklus i .

[0029] Ein Funktionsblock 7 bildet fortlaufend aus den Abtastwerten des Sprachsignals $x2(i)$ Schätzwerte $\text{SNR}(i)$ des zugehörigen Signal-/ Rauschleistungsverhältnisses, die von einem Funktionsblock 8 ausgewertet werden. Auch eine Auswertung des Sprachsignals $x1(i)$ anstelle des Sprachsignals $x2(i)$ ist möglich, ohne daß die Funktionsfähigkeit der Sprachverarbeitungsvorrichtung eingeschränkt wird. Die Funktionsweise des Funktionsblockes 7 wird später anhand der Figuren 6 bis 8 näher erläutert. Der Funktionsblock 8 führt eine Schwellwertentscheidung bezüglich der Schätzwerte $\text{SNR}(i)$ durch. Nur wenn die Schätzwerte $\text{SNR}(i)$ über einer vorgebbaren Schwelle liegen, wird ein Zwischenspeicher 9 mit dem neu bestimmten Gradientenschätzwert $\text{grad}(i)$ überschrieben. Dieser Fall wird durch die geschlossene Stellung eines Schalters 11 symbolisiert, der von dem Funktionsblock 8 gesteuert wird. Der Speicherinhalt ($\text{grad}(i)$) des Zwischenspeichers 9 wird von einer Funktionseinheit 10 weiterverarbeitet. Für den Fall, daß ein Schätzwert $\text{SNR}(i)$ unterhalb des vorgebbaren Schwellwerts liegt, wird der Zwischenspeicher 9 nicht mit dem neu ermittelten Gradientenschätzwert $\text{grad}(i)$ überschrieben und er behält seinen alten Speicherinhalt bei, was durch die geöffnete Stellung des Schalters 11 symbolisiert wird. Die vorgebbare Schwelle, von der das Öffnen und Schließen des Schalters 11 durch den Funktionsblock 8 abhängt, liegt vorzugsweise zwischen 0 und 10 dB.

[0030] Der Zwischenspeicher 9 liefert die in ihm gespeicherten Gradientenschätzwerte $\text{grad}(i)$ an die Funktionseinheit 10, der auch Abtastwerte des Sprachsignals $x1(i)$ zugeführt werden und die sowohl zur Lieferung der Sprachsignalschätzwerte $x1_{\text{int}}(i)$ als auch zur Einstellung des Verzögerungsgliedes 4 dient.

[0031] Die Gradientenschätzwerte $\text{grad}(i)$ werden von einem Funktionsblock 12 nach

$$\text{sgrad}(i) = \alpha * \text{sgrad}(i-1) + (1-\alpha) * \text{grad}(i) \quad (5)$$

zu geglätteten ("smoothed") Gradientenschätzwerten $\text{sgrad}(i)$ weiterverarbeitet. α ist eine Konstante, die im Ausführungsbeispiel den Wert 0,95 besitzt. Die Werte $\text{sgrad}(i)$ werden von einem Funktionsblock 13 zur Adaption von Verzögerungsschätzwerten $T1'(i)$ nach

$$T1'(i+1) = T1'(i) - \mu * \text{sgrad}(i) \quad (6)$$

verwendet. Die Bestimmung von Verzögerungsschätzwerten $T1'(i)$ erfolgt damit rekursiv. μ ist ein konstanter Faktor bzw. Konvergenzparameter und liegt im Bereich

$$0 < \mu < \frac{1}{10 * R_{x2x2}} \quad (7)$$

R_{x2x2} bezeichnet eine Autokorrelationsfunktion des Sprachsignals $x2(i)$ an der Stelle Null. Ein besonders vorteilhafter Wertebereich von μ ist im vorliegenden Ausführungsbeispiel $1,5 < \mu < 3$.

[0032] Die Verzögerungsschätzwerte $T1'(i)$ können auch nicht ganzzahlige Werte d.h. nicht ganzzahlige Vielfache eines Abtastintervalls sein. Ein Funktionsblock 14 rundet die Verzögerungsschätzwerte $T1'(i)$ auf ganzzahlige Verzögerungswerte $T1(i)$, mit denen die Verzögerungsvorrichtung 4 eingestellt wird. Die Rundungsoperation durch Funktionsblock 14 ist notwendig, da Werte des durch das Verzögerungsglied 4 zu verzögernden Sprachsignals $x1(i)$ nur zu den entsprechenden Abtastzeitpunkten vorliegen.

[0033] Die Funktionseinheit 10 weist weiterhin einen Funktionsblock 15 auf, der die Sprachsignalschätzwerte $x1_{\text{int}}(i)$ nach

$$x1_{\text{int}}(i) = x1(i+T1(i)) + 0,5 * [T1'(i) - T1(i)] * [x1(i+T1(i)+1)) - x1(i+T1(i)-1))] \quad (8)$$

durch Interpolation dreier benachbarter Abtastwerte $x1(i+T1(i)-1)$, $x1(i+T1(i))$ und $x1(i+T1(i)+1)$ des Sprachsignals $x1$ bildet. Der Funktionsblock 15 ist somit in der Lage, durch den Sprachsignalschätzwert $x1_{\text{int}}(i)$ im Programmzyklus i einen Wert des Sprachsignals $x1$ zum Zeitpunkt $i+T1(i)$, d.h. zu einem Zeitpunkt zwischen zwei Abtastzeitpunkten, zu bilden bzw. zu interpolieren. Die beschriebene Interpolation durch Funktionsblock 15 kann dadurch ersetzt werden, daß Funktionsblock 15 eine Tiefpaßfilterung der Abtastwerte $x1(i)$ zur Interpolation von Werten zwischen den Abtastzeitpunkten durchführt.

[0034] Würden zur Bestimmung der Fehlerwerte $e_{12}(i)$ anstelle der Sprachsignalschätzwerte $x_{1\text{int}}(i)$ die am Ausgang des Verzögerungsgliedes 4 anliegenden verzögerten Abtastwerte des Sprachsignals $x_1(i)$ verwendet, wie dies aus "IEEE Transactions on Acoustics, Speech, and Signal Processing, VOL. ASSP-29, Nr.3, Juni 1981, S. 582-587" bekannt ist, würde beim Erreichen von Fehlerwerten $e_{12}(i) = 0$ die Verzögerungswerte $T_1(i)$, mit denen das Verzögerungsglied 4 eingestellt wird, nicht mehr konvergieren. Es ergäben sich starke Oszillationen der gerundeten Verzögerungswerte $T_1(i)$. Diese würden zwischen zwei Verzögerungswerten mit dem Abstand eines Abtastintervalls schwanken. Die entsprechende wahre Zeitverzögerung zwischen den Sprachsignalanteilen, die durch die unterschiedlichen Wegstrecken vom Sprecher zu den Mikrofonen M1 und M2 bestimmt ist, würde dabei zwischen diesen zwei Verzögerungswerten liegen. Im vorliegenden Ausführungsbeispiel werden solche Oszillationen dadurch vermieden, daß bei der Bildung der Fehlerwerte Sprachsignalschätzwerte $x_{1\text{int}}(i)$ verwendet werden, durch die die Werte des Sprachsignals $x_1(i)$ auch für Verzögerungen um nicht ganzzahlige Vielfache eines Abtastintervalls verfügbar sind, also auch an Zeitpunkten ungleich der Abtastzeitpunkte i des Sprachsignals $x_1(i)$.

[0035] Der zur Glättung der Gradientenschätzwerte $\text{grad}(i)$ dienende Funktionsblock 12 bewirkt eine verbesserte Ermittlung der Verzögerungsschätzwerte $T_1'(i)$.

[0036] Die Steuervorrichtung 3 adaptiert die Verzögerungsschätzwerte $T_1'(i)$ bzw. die Verzögerungswerte $T_1(i)$ so, daß von einem Programmzyklus zum nächsten das Quadrat bzw. die Leistung der Fehlerwerte $e_{12}(i)$ verringert wird. Die Konvergenz von $T_1'(i)$ bzw. $T_1(i)$ ist somit sichergestellt.

[0037] In Fig. 3 ist eine prinzipiell wie die Sprachverarbeitungsvorrichtung aus Fig. 1 arbeitende Sprachverarbeitungsvorrichtung mit nun drei Mikrofonen M1, M2 und M3 zur Lieferung von Mikrofon- bzw. Sprachsignalen dargestellt. Die Mikrophonsignale werden Analog-Digital-Umsetzern 20, 21 und 22 zugeführt, die digitalisierte und damit abgetastete Sprachsignale $x_1(i)$, $x_2(i)$ und $x_3(i)$ liefern, die aus Sprach- und Rauschsignalanteilen bestehen. Die Sprachsignale $x_1(i)$ und $x_3(i)$ werden einstellbaren Verzögerungsgliedern 23 und 24 zugeführt. Analog zu Fig. 1 wird das Sprachsignal $x_2(i)$ einem Verzögerungsglied 27 mit einer festen Verzögerungszeit T_{max} zugeführt. Die Ausgangswerte der Verzögerungsglieder 23, 24 und 27 werden von einer Addiervorrichtung 25 zum Summensignal $X(i)$ aufaddiert. Eine Steuervorrichtung 26 wertet die Abtastwerte der Sprachsignale $x_1(i)$, $x_2(i)$ und $x_3(i)$ aus und leitet aus diesen Abtastwerten analog zur Wirkungsweise der Steuervorrichtung 3 aus Fig. 1 und 2 gerundete ganzzahlige Verzögerungswerte $T_1(i)$ und $T_3(i)$ ab, die ganzzahligen Vielfachen eines Abtastintervalls der abgetasteten Sprachsignale $x_1(i)$, $x_2(i)$ und $x_3(i)$ entsprechen und mit denen die Verzögerungsglieder 23 und 24 eingestellt werden, so daß eine Erweiterung von zwei auf drei zu verarbeitende Mikrofon- bzw. Sprachsignale ermöglicht wird.

[0038] In Fig. 4 ist eine erste Ausführungsform der Steuervorrichtung 26 aus Fig. 3 dargestellt. Es sind zwei Funktionseinheiten 10 vorgesehen, deren Aufbau gleich dem Aufbau der Funktionseinheit 10 aus Fig. 2 ist und die zur Einstellung der Verzögerungsglieder 23 und 24 mit den gerundeten Zeitverzögerungswerten $T_1(i)$ und $T_3(i)$ dienen.

[0039] Die obere Funktionseinheit 10 liefert Sprachsignalschätzwerte $x_{1\text{int}}(i)$. Die untere Funktionseinheit 10 liefert Sprachsignalschätzwerte $x_{3\text{int}}(i)$. Aus einer Differenz $x_{1\text{int}}(i) - x_2(i)$ und aus einer Differenz $x_{3\text{int}}(i) - x_2(i)$ werden Fehlerwerte $e_{12}(i)$ und $e_{32}(i)$ gebildet.

[0040] Auch hier ist ein Digitalfilter 6 vorgesehen, das in den Ausführungen zu Fig. 2 bereits näher beschrieben ist, und das zum Empfang der Abtastwerte $x_2(i)$ und zur Lieferung von Werten $x_{2H}(i)$ dient, die durch eine Hilbert-Transformation der Abtastwerte $x_2(i)$ erzeugt werden. Die Werte $x_{2H}(i)$ werden einerseits mit den Fehlerwerten $e_{12}(i)$ und andererseits mit den Fehlerwerten $e_{32}(i)$ multipliziert. Das erste Produkt $x_{2H}(i) \cdot e_{12}(i)$ wird der oberen, das zweite Produkt $x_{2H}(i) \cdot e_{32}(i)$ wird der unteren Funktionseinheit 10 zugeführt. Die Anordnung der Funktionsblöcke 7 und 8, des Zwischenspeichers 9 und des Schalters 11 wird analog zu Fig. 2 durchgeführt und ist aus Gründen der Übersichtlichkeit nicht in Fig. 4 dargestellt.

[0041] Fig. 5 zeigt eine gegenüber Fig. 4 erweiterte Fassung der Steuervorrichtung 26. Im Gegensatz zu Fig. 4 sind anstelle nur eines Digitalfilters 6 nun drei Digitalfilter 6 angeordnet. Diese bilden aus den Sprachsignalabtastwerten $x_1(i)$, $x_2(i)$ und $x_3(i)$ durch Hilbert-Transformation die Werte $x_{1H}(i)$, $x_{2H}(i)$ und $x_{3H}(i)$.

[0042] In der oberen Hälfte des in Fig. 5 dargestellten Blockdiagramms werden Fehlerwerte $e_{13}(i)$ aus der Differenz $x_{1\text{int}}(i) - x_2(i)$ gebildet, die in ein erstes Produkt $0,3 \cdot e_{13}(i) \cdot x_{3H}(i)$ eingehen. Ein zweites Produkt ergibt sich aus $0,7 \cdot e_{12}(i) \cdot x_{2H}(i)$. Die beiden Produkte entsprechen gewichteten Gradientenschätzwerten der Quadrate der Fehlerwerte $e_{13}(i)$ und $e_{12}(i)$. Die Summe aus erstem und zweitem Produkt und damit eine Linearkombination der gewichteten Gradientenschätzwerten wird der oberen Funktionseinheit 10 zugeführt.

[0043] Analog dazu werden in der unteren Hälfte des in Fig. 5 dargestellten Blockdiagramms Fehlerwerte $e_{31}(i)$ und $e_{32}(i)$ gebildet. Die Fehlerwerte $e_{31}(i)$ ergeben sich aus der Differenz $x_{3\text{int}}(i) - x_1(i)$. Die Fehlerwerte $e_{32}(i)$ werden durch die Differenz $x_{3\text{int}}(i) - x_2(i)$ gebildet. Ein drittes Produkt $0,3 \cdot e_{31}(i) \cdot x_{1H}(i)$ und ein viertes Produkt $0,7 \cdot e_{32}(i) \cdot x_{2H}(i)$ werden aufaddiert und die sich ergebende Summe wird der unteren Funktionseinheit 10 zugeführt.

[0044] Mit Hilfe der Sprachverarbeitungsvorrichtung nach Fig. 3, die eine Steuervorrichtung nach Fig. 4 oder 5 enthält, läßt sich ein gegenüber der Sprachverarbeitungsvorrichtung mit zwei Mikrofonen nach Fig. 1 verbessertes Summensignal $X(i)$ erzeugen. Das Signal-/ Rauschleistungsverhältnis und damit die Sprachqualität des Summensignals $X(i)$ der Sprachverarbeitungsvorrichtung nach Fig. 3 ist gegenüber dem von der Sprachverarbeitungsvorrichtung nach

Fig. 1 erzeugten Summensignal $X(i)$ weiter erhöht. Die Steuervorrichtung nach Fig. 5 weist gegenüber der Steuervorrichtung nach Fig. 4 beim Einsatz in der Sprachverarbeitungsvorrichtung nach Fig. 3 eine erhöhte Stabilität auf.

[0045] Sowohl in Fig. 4 als auch in Fig. 5 ist aus Gründen der Übersichtlichkeit auf eine Darstellung von Mitteln (siehe Funktionsblöcke 7 und 8, Zwischenspeicher 9 und Schalter 11 in Fig. 2) verzichtet worden, die eine Abhängigkeit der Sprachverarbeitung von Schätzwerten $SNR(i)$ für eines der Mikrophonsignale $x_1(i)$, $x_2(i)$ oder $x_3(i)$ bewirken. Ebenfalls aus Gründen der Übersichtlichkeit ist die Normierung von Produkten aus Fehlerwerten und der Ausgangswerte der die Hilbert-Transformation durchführenden Digitalfilter 6 auf die Leistung eines zugehörigen Mikrophonsignals (siehe $1/P_{x_2}(i)$ in Fig. 2) nicht dargestellt. Die Erweiterung der Steuervorrichtungen 26 nach Fig. 4 und 5 um diese beiden technischen Merkmale ergibt sich aus ihrer Realisierung in der Steuervorrichtung 3 nach Fig. 2.

[0046] Zur Erhöhung der Sprachqualität der Summensignale $X(i)$ am Ausgang der Addiervorrichtungen 5 und 25 in Fig. 1 und Fig. 3 kann die Erfindung so ausgestaltet werden, daß die Verzögerungsschätzwerte $T1'(i)$ und $T3'(i)$ (das sind z. B. Fließkommazahlen) zur Bildung der Verzögerungswerte $T1(i)$ und $T3(i)$ nicht auf Werte gerundet werden, die einem ganzzahligen Vielfachen eines Abtastintervalls entsprechen (hier: ganze Zahlen), sondern auf Werte, die einem Vielfachen eines Bruchteils eines Abtastintervalls entsprechen. Insbesondere ist eine Rundung der Verzögerungsschätzwerte auf Vielfache eines Wertes vorteilhaft, der einem Viertel oder der Hälfte eines Abtastintervalls entspricht. Auf diese Weise wird die Auflösung der Verzögerungswerte erhöht, die somit genauer einstellbar sind, so daß auch die Sprachqualität der Summensignale $X(i)$ weiter erhöht wird, da Laufzeitunterschiede von der die Sprachsignalanteile erzeugenden Sprachquelle zu den Mikrofonen M1, M2 und M3 genauer ausgeglichen werden können. Bei der Verzögerung eines Sprachsignals mit einem Vielfachen eines Bruchteils eines Abtastintervalls wird eine Interpolation oder Tiefpaßfilterung von Sprachsignalabtastwerten vorgesehen, um Sprachsignalwerte zu erzeugen, die zwischen jeweils zwei Sprachsignalabtastwerten liegen. Die Interpolation bzw. Tiefpaßfilterung kann insbesondere in die Verzögerungsmittel 4, 23 und 24 integriert werden.

[0047] Mit Hilfe der Fig. 6 und 7 wird das Schema erläutert, anhand dessen der Funktionsblock 7 aus einem abgetasteten Sprachsignal $x(i)$, das aus Rausch- und Sprachsignalanteilen besteht, die zugehörigen Schätzwerte $SNR(i)$ des Signal-/ Rauschleistungsverhältnisses, d.h. des Verhältnisses der Leistungen der Sprachsignalanteile zur Leistung der Rauschsignalanteile, ermittelt. Den Abtastwerten $x(i)$ entsprechen in Fig. 2 die Abtastwerte $x_2(i)$. In Fig. 6 ist der Funktionsblock 7 anhand eines Blockschaltbildes dargestellt. Ein Funktionsblock 30 dient zur Bildung von Leistungswerten $P_x(i)$ der Abtastwerte $x(i)$ durch Quadrieren der Abtastwerte. Weiterhin führt der Funktionsblock 30 eine Glättung dieser Leistungswerte $P_x(i)$ durch. Die sich so ergebenden geglätteten Leistungswerte $P_{x,s}(i)$ werden sowohl dem Funktionsblock 31 als auch dem Funktionsblock 32 zugeführt. Der Funktionsblock 31 ermittelt fortlaufend Schätzwerte $P_n(i)$ zur Abschätzung der Leistung des Rauschsignalanteils der Abtastwerte $x(i)$, d.h. es wird die Leistung der Rauschsignalanteile der Abtastwerte $x(i)$ ermittelt. Aus den geglätteten Leistungswerten $P_{x,s}(i)$ und den Schätzwerten $P_n(i)$ bestimmt der Funktionsblock 32 fortlaufend Schätzwerte $SNR(i)$ des Signal-/ Rauschleistungsverhältnisses der Abtastwerte $x(i)$.

[0048] In Fig. 7 ist ein Flußdiagramm dargestellt, das die Funktionsweise des Funktionsblockes 7 näher erläutert. Anhand des Flußdiagramms wird ersichtlich, wie aus den Abtastwerten $x(i)$ des Sprachsignals x durch ein Computerprogramm Schätzwerte $SNR(i)$ des entsprechenden Signal-/ Rauschleistungsverhältnisses gebildet werden. In einem Initialisierungsblock 33 wird zu Beginn des durch Fig. 7 beschriebenen Programms eine Zählvariable Z auf 0 und eine Variable P_{Mmin} auf einen Wert P_{max} gesetzt. P_{max} ist so groß gewählt, daß die geglätteten Leistungswerte $P_{x,s}(i)$ immer kleiner als P_{max} sind. P_{max} kann beispielsweise auf den maximal darstellbaren Zahlenwert eines zur Realisierung des Programms verwendeten Rechners gesetzt werden. In einem Block 34 wird ein neuer Abtastwert $x(i)$ eingelesen. In Block 35 wird eine Zählvariable Z um den Wert 1 erhöht, wonach in Block 36 ein neuer geglätteter Leistungswert $P_{x,s}(i)$ gebildet wird. Er ergibt sich dadurch, daß zunächst durch

$$P_x(i) = P_x(i-1) + x^2(i) - x^2(i-N) \quad (9)$$

ein Kurzzeitleistungswert $P_x(i)$ und dann durch

$$P_{x,s}(i) = \alpha * P_{x,s}(i-1) + (1-\alpha) * P_x(i) \quad (10)$$

ein neuer geglätteter Leistungswert gebildet wird. Mit Formel (9) wird ein Kurzzeitleistungswert $P_x(i)$ einer Gruppe von N aufeinanderfolgenden Abtastwerten $x(i)$ ermittelt. N ist hier beispielsweise gleich 128. Der Wert α aus Gleichung (10) liegt zwischen 0,95 und 0,98. Die Ermittlung von geglätteten Leistungswerten $P_{x,s}(i)$ kann auch nur durch Gleichung (10) durchgeführt werden, wobei dann allerdings der Wert α ungefähr auf den Wert 0,99 zu erhöhen und $P_x(i)$ durch $x^2(i)$ zu ersetzen ist.

[0049] Durch eine Verzweigung 37 wird danach abgefragt, ob der gerade ermittelte geglättete Leistungswert $P_{x,s}(i)$ kleiner als P_{Mmin} ist. Wird diese Frage bejaht, d.h. $P_{x,s}(i)$ ist kleiner als P_{Mmin} , wird durch Block 38 P_{Mmin} auf den Wert von $P_{x,s}(i)$ gesetzt. Falls die Frage von Verzweigung 37 verneint wird, wird Block 38 übersprungen. Damit steht in P_{Mmin} nach M Programmzyklen das Minimum von M geglätteten Leistungswerten $P_{x,s}$. Danach erfolgt mit der Verzweigung 39 die Abfrage, ob die Zählervariable Z einen Wert größer oder gleich einem Wert M hat. Es wird auf diese Weise festgestellt, ob schon M geglättete Leistungswerte abgearbeitet sind.

[0050] Wird die Frage von Verzweigung 39 verneint, d.h. es sind noch nicht M geglättete Leistungswerte abgearbeitet, wird das Programm mit Block 40 fortgesetzt. Dort wird ein vorläufiger Schätzwert $P_n(i)$ der Rauschsignalleistung des Sprachsignals x durch

$$P_n(i) = \min \{P_{x,s}(i), P_n(i)\} \quad (11)$$

bestimmt. Diese Operation stellt sicher, daß der vorläufige Schätzwert $P_n(i)$ nicht größer als der aktuelle geglättete Leistungswert $P_{x,s}(i)$ sein kann. Danach wird mit Block 41 nach der Formel

$$SNR(i) = [P_{x,s}(i) - \min\{c \cdot P_n(i), P_{x,s}(i)\}] / [c \cdot P_n(i)] \quad (12)$$

ein aktueller Schätzwert $SNR(i)$ des Signal-/Rauschleistungsverhältnisses des Sprachsignals x(i) ermittelt. Im Normalfall dient das Produkt $c \cdot P_n(i)$ zur Abschätzung der aktuellen Leistung des Rauschsignalanteils, und die Differenz $P_{x,s}(i) - c \cdot P_n(i)$ dient zur Abschätzung der aktuellen Leistung des Sprachsignalanteils des Sprachsignals x(i). Die aktuelle Leistung des Sprachsignals wird durch den geglätteten Leistungswert $P_{x,s}(i)$ geschätzt. Die Gewichtung mit einem Skalierungsfaktor c verhindert, daß durch $P_n(i)$ die Rauschsignalleistung mit einem zu kleinen Wert abgeschätzt wird. Der Skalierungsfaktor c liegt typisch im Bereich von 1,3 bis 2. Durch die Minimumbildung in Block 41 bzw. Gleichung (12) wird sichergestellt, daß das nicht logarithmierte Signal-/ Rauschleistungsverhältnis $SNR(i)$ auch dann positiv ist, wenn im Ausnahmefall $c \cdot P_n(i)$ größer als $P_{x,s}(i)$ ist. Dann wird die Leistung des Rauschsignalanteils des Sprachsignals gleich der durch $P_{x,s}(i)$ geschätzten Leistung des Sprachsignals gesetzt. Die durch $P_{x,s}(i) - P_{x,s}(i)$ geschätzte Leistung des Sprachsignalanteils des Sprachsignals ist dann wie auch das nicht logarithmische Signal-/ Rauschleistungsverhältnis gleich Null. Das Programm wird nach der Berechnung des Schätzwertes $SNR(i)$ mit dem Einlesen eines neuen Sprachsignalabtastwertes x(i) durch Block 34 fortgesetzt.

[0051] Wird die Abfrage von Verzweigung 39 bejaht, d.h. es sind M geglättete Abtastwerte $P_{x,s}(i)$ abgearbeitet, werden in Block 42 durch

$$\begin{aligned} \text{minvec}_1 &= \text{minvec}_2; \\ \text{minvec}_2 &= \text{minvec}_3; \\ &: \\ \text{minvec}_{W-1} &= \text{minvec}_W; \\ \text{minvec}_W &= P_{Mmin}; \end{aligned} \quad (13)$$

die Komponenten eines Vektors minvec der Dimension W aktualisiert. Danach wird durch Verzweigung 43 abgefragt, ob die Komponenten minvec_1 bis minvec_W mit ansteigendem Vektorindex ansteigen, d.h. ob gilt:

$$\text{minvec}_{j+1} > \text{minvec}_j \text{ für } 1 \leq j \leq W-1 \quad (14)$$

[0052] Wird die Abfrage von Verzweigung 43 verneint, d.h. die zuletzt ermittelten in den Komponenten des Vektors minvec stehenden zuletzt ermittelten W Minima steigen nicht monoton an, wird durch Block 44 nach

$$P_n(i) = \min\{\text{minvec}_W, \text{minvec}_{W-1}, \dots, \text{minvec}_1\} \quad (15)$$

der vorläufige Schätzwert $P_n(i)$ der Rauschsignalleistung aus den Minima der Komponenten des Vektors minvec , d.h. aus dem Minimum der letzten $L=W \cdot M$ aufeinanderfolgenden geglätteten Leistungswerte $P_{x,s}(i)$, bestimmt. Bei einer Bejahung der durch Verzweigung 43 gestellten Frage, d.h. bei einem monotonen Ansteigen der zuletzt ermittelten in den Komponenten des Vektors minvec stehenden W Minima wird in Block 45 $P_n(i)$ gleich P_{Mmin} gesetzt, so daß eine Anpassung der Abschätzung des Rauschsignalanteils beschleunigt erfolgt, da $P_n(i)$ an dem Minimum des letzten ($M < L$) Werte bestimmt wird. Danach wird in Block 46 die Zählervariable Z wieder auf 0 gesetzt und P_{Mmin} erhält erneut den Wert P_{max} .

[0053] Durch das beschriebene Programm werden jeweils M aufeinanderfolgende geglättete $P_{x,s}(i)$ Abtastwerte $x(i)$ des Sprachsignals x zu einer Untergruppe zusammengefaßt. Innerhalb einer solchen Untergruppe wird durch die mit Verzweigung 37 und Block 38 durchgeführten Operationen das Minimum der geglätteten Leistungswerte $P_{x,s}(i)$ ermittelt. Die zuletzt ermittelten W Minima werden in den Komponenten des Vektors minvec abgespeichert. Sind die letzten W Minima nicht monoton ansteigend (siehe Verzweigung 43), so wird nach Block 44 ein vorläufiger Schätzwert $P_n(i)$ der Leistung des Rauschsignalanteils aus dem Minimum der Minima der letzten W Untergruppen, d.h. aus dem Minimum einer Gruppe, bestimmt. Es werden jeweils zur Bildung einer Gruppe mit $L=W \cdot M$ aufeinanderfolgenden geglätteten Leistungswerten $P_{x,s}(i)$ W aufeinanderfolgende Untergruppen zusammengefaßt. Die Gruppen mit jeweils L Werten folgen lückenlos aufeinander und überlappen sich jeweils mit $L-M$ geglätteten Leistungen $P_{x,s}(i)$.

[0054] Für den Fall, daß die Minima von W aufeinanderfolgenden Untergruppen monoton ansteigen (siehe Verzweigung 43), wird durch Block 45 zur Abschätzung des aktuellen Schätzwertes $P_n(i)$ der Leistung des Rauschsignalanteils jeweils das Minimum der letzten Untergruppe mit M geglätteten Leistungswerten $P_{x,s}(i)$ verwendet. Die Zeitspanne, mit der monoton ansteigende geglättete Leistungswerte $P_{x,s}(i)$ auch eine Änderung der Schätzwerte $\text{SNR}(i)$ bewirken, wird damit verkürzt.

[0055] Fig. 8 verdeutlicht, wie die geglätteten Leistungswerte $P_{x,s}$ in Gruppen und Untergruppen zusammengefaßt werden. Es werden jeweils M geglättete Leistungswerte $P_{x,s}(i)$, die jeweils zu Abtastzeitpunkten i vorliegen, zu einer Untergruppe zusammengefaßt. Die Untergruppen grenzen aneinander. Für jede Untergruppe wird das Minimum der geglätteten Leistungswerte $P_{x,s}(i)$ bestimmt. Jeweils W Untergruppenminima werden in dem Vektor minvec abgespeichert. In der Regel, d.h. bei nicht monoton ansteigenden W Untergruppen Minima, werden W Untergruppen zu einer Gruppe mit $L = W \cdot M$ geglätteten Leistungswerten $P_{x,s}(i)$ zusammengefaßt. Nach jeweils M geglätteten Leistungen $P_{x,s}(i)$ wird aus dem Minimum der letzten W Untergruppenminima bzw. der letzten L geglätteten Leistungswerte $P_{x,s}(i)$ der Wert $P_n(i)$ bestimmt, der zur Abschätzung der Rauschsignalleistung dient. In Fig. 8 sind acht Gruppen mit jeweils L Abtastwerten $x(i)$ dargestellt, die jeweils $W = 4$ Untergruppen mit M geglätteten Leistungswerten $P_{x,s}(i)$ enthalten. Die acht Gruppen überlappen sich teilweise. So enthalten zwei aufeinanderfolgende Gruppen jeweils $L-M$ gleiche geglättete Leistungswerte $P_{x,s}(i)$. Auf diese Weise wird ein guter Kompromiß zwischen dem erforderlichen Rechenaufwand und der jeweiligen Verzögerungszeit erreicht, mit der eine Aktualisierung eines Schätzwertes $P_n(i)$ der Rauschsignalleistung zur Aktualisierung eines Schätzwertes $\text{SNR}(i)$ des Signal/ Rauschleistungsverhältnisses erfolgt. Eine Realisierung mit aneinandergrenzenden, d.h. sich nicht überlappenden Gruppen ist auch denkbar. Allerdings ist dann bei verringertem Rechenaufwand die Zeitspanne zwischen zwei Schätzwerten $\text{SNR}(i)$ vergrößert, so daß die Reaktionszeit auf sich ändernde SNR des Sprachsignals $x(i)$ vergrößert ist.

[0056] Die beschriebene Sprachverarbeitungsvorrichtung weist damit eine Schätzvorrichtung auf, die zum fortlaufenden Bilden von Schätzwerten $\text{SNR}(i)$ des Signal-/ Rauschleistungsverhältnisses von verrauschten Sprachsignalen $x(i)$ geeignet ist. Insbesondere sind keine Sprachpausen zur Abschätzung der Rauschsignalleistung erforderlich. Die beschriebene Schätzvorrichtung nutzt den besonderen Zeitverlauf von geglätteten Leistungswerten des Sprachsignals $x(i)$ aus, der durch Spitzen und dazwischenliegende Bereiche mit kleineren geglätteten Leistungswerten $P_{x,s}(i)$ gekennzeichnet ist, deren zeitliche Ausdehnung von der jeweiligen Sprachquelle, d.h. dem jeweiligen Sprecher, abhängt. Dabei werden die Bereiche zwischen den Spitzen zur Abschätzung der Leistung des Rauschsignalanteils verwendet. Die Gruppen mit jeweils L geglätteten Leistungswerten $P_{x,s}(i)$ müssen lückenlos aufeinanderfolgen, d.h. sie müssen entweder aneinandergrenzen oder sich überlappen. Weiterhin muß sichergestellt sein, daß mindestens ein Wert eines zwischen zwei Spitzen liegenden Bereichs mit kleineren geglätteten Leistungswerten $P_{x,s}(i)$ von jeder Gruppe erfaßbar ist, d.h. jede Gruppe muß so viele geglättete Leistungswerte $P_{x,s}(i)$ enthalten, daß mindestens alle zu einer beliebigen Spitze gehörenden Werte erfaßbar sind. Da die zeitlich ausgedehntesten Spitzen jeweils durch die zeitlich ausgedehntesten Phoneme eines Sprachsignals, d.h. die Vokale, abschätzbar sind, kann daraus die die Gruppengröße beschreibende Zahl L abgeleitet werden. Für eine Abtastrate des Sprachsignals von 8 kHz liegt ein sinnvoller Wert von L im Bereich zwischen 3000 und 8000. Ein vorteilhafter Wert für W ist 4. Bei einer solchen Dimensionierung ergibt sich ein guter Kompromiß zwischen Rechenaufwand und Reaktionsschnelligkeit des Funktionsblockes 7.

[0057] In Fig. 9 ist eine Verwendung der Sprachverarbeitungsvorrichtung aus Fig. 3 in einem Mobilfunkendgerät 50 dargestellt. Die Sprachverarbeitungsmittel 20 bis 26 sind in einem Funktionsblock 51 zusammengefaßt, der aus den von den Mikrofonen $M1$, $M2$ und $M3$ erzeugten Mikrophon- bzw. Sprachsignalen die Summensignalwerte $X(i)$ bildet. Die Mikrophone $M1$, $M2$ und $M3$ haben vorteilhaft einen Abstand von 10 bis 60 cm, so daß in einer sogenannten "verhallten" Umgebung (z.B. Auto, Büro) die Störsignalanteile der von den Mikrofonen $M1$, $M2$ und $M3$ gelieferten

Sprachsignale weitgehend unkorreliert sind. Dies gilt auch beim Einsatz von nur zwei Mikrofonen wie in Fig. 1. Ein die Summensignalwerte $X(i)$ verarbeitender Funktionsblock 52 faßt alle übrigen Mittel des Mobilfunkendgerätes 50 zum Empfang, Verarbeiten und Senden von Signalen zusammen, welche zur Kommunikation mit einer nicht dargestellten Basisstation dienen, wobei das Senden und Empfangen von Signalen über eine an den Funktionsblock 52 gekoppelte Antenne 54 erfolgt. Weiterhin ist ein mit dem Funktionsblock 52 gekoppelter Lautsprecher 53 vorgesehen. Die akustische Kommunikation eines Benutzers (Sprecher, Hörer) mit dem Mobilfunkendgerät 50 erfolgt über die Mikrophone M1 bis M3 und den Lautsprecher 53, die Teile einer in das Mobilfunkendgerät 50 integrierten Freisprecheinrichtung sind. Die Anwendung eines solchen Mobilfunkendgerätes 50 ist insbesondere in Kraftfahrzeugen von Vorteil, da dort das Freisprechen über das Mobilfunkendgerät insbesondere durch Motor- oder Fahrgeräusche (Rauschen) gestört ist.

Patentansprüche

1. Sprachverarbeitungsvorrichtung zur Verarbeitung eines ersten ($x_2(i)$) und mindestens eines weiteren ($x_1(i)$, $x_3(i)$) aus Rausch- und Sprachsignalanteilen bestehenden und als Abtastwerte vorliegenden Sprachsignals mit Verzögerungsmitteln (4, 23, 24) zur Verzögerung des abgetasteten weiteren Sprachsignals ($x_1(i)$, $x_3(i)$), mit Steuermitteln (3, 26)
 - zur Bildung von Gradientenschätzwerten ($\text{grad}(i)$, $\text{sgrad}(i)$) durch Multiplikation von Fehlerwerten ($e_{12}(i)$, $e_{32}(i)$, $e_{13}(i)$, $e_{31}(i)$) für zwei Sprachsignale (z.B. $x_1(i)$ und $x_2(i)$) mit den Ausgangswerten eines Digitalfilters (6), das eine Phasenverschiebung von 90 Grad bewirkt und zur Filterung eines der zwei Sprachsignale (z.B. $x_2(i)$) dient,
 - zur rekursiven Ermittlung von Verzögerungsschätzwerten ($T_1'(i)$, $T_3'(i)$) aus den Gradientenschätzwerten ($\text{grad}(i)$, $\text{sgrad}(i)$), wobei aus den Verzögerungsschätzwerten ($T_1'(i)$, $T_3'(i)$) durch Rundung die Verzögerungswerte ($T_2(i)$, $T_3(i)$) gebildet werden, die zur Einstellung der Verzögerungsmittel (4, 23, 24) dienen und
 - zur Bildung jeweils wenigstens eines Fehlerwertes ($e_{12}(i)$, $e_{32}(i)$, $e_{13}(i)$, $e_{31}(i)$) für einen bestimmten Abtastzeitpunkt (i) aus der Differenz zwischen einem Sprachsignalschätzwert ($x_{1\text{int}}(i)$, $x_{3\text{int}}(i)$), der zur Abschätzung des weiteren Sprachsignals ($x_1(i)$, $x_3(i)$) zu einem gegenüber dem bestimmten Abtastzeitpunkt (i) um den Verzögerungsschätzwert ($T_1'(i)$, $T_3'(i)$) verschobenen Zeitpunkt dient und durch Interpolation von Abtastwerten des weiteren Sprachsignals ($x_1(i)$, $x_3(i)$) gebildet wird, und dem Abtastwert eines anderen der zu verarbeitenden Sprachsignale ($x_1(i)$, $x_2(i)$, $x_3(i)$) zu dem bestimmten Abtastzeitpunkt (i) und mit einer Addiervorrichtung (5, 25) zum Addieren der gegeneinander zeitversetzten Sprachsignale ($x_1(i)$, $x_2(i)$, $x_3(i)$).
2. Mobilfunkendgerät mit einer Sprachverarbeitungsvorrichtung nach Anspruch 1.
3. Mobilfunkendgerät nach Anspruch 2, dadurch gekennzeichnet, daß das Digitalfilter (6) ein digitaler Hilbert-Transformator ist.
4. Mobilfunkendgerät nach Anspruch 3, dadurch gekennzeichnet, daß Mittel (12) zur Glättung der Gradientenschätzwerte ($\text{grad}(i)$) vorgesehen sind.
5. Mobilfunkendgerät nach einem der Ansprüche 2 bis 4, dadurch gekennzeichnet, daß die Sprachverarbeitungsvorrichtung zur Verarbeitung von drei Sprachsignalen ($x_1(i)$, $x_2(i)$, $x_3(i)$) vorgesehen ist.
6. Mobilfunkendgerät nach einem der Ansprüche 2 bis 5, dadurch gekennzeichnet, daß zur Ermittlung eines Verzögerungsschätzwertes ($T_1'(i)$, $T_3'(i)$) für das weitere Sprachsignal ($x_1(i)$, $x_3(i)$) die Verwendung einer Linearkombination von Fehlerwerten ($e_{12}(i)$ mit $e_{13}(i)$, $e_{31}(i)$ mit $e_{32}(i)$) vorgesehen ist.
7. Mobilfunkendgerät nach einem der Ansprüche 2 bis 6, dadurch gekennzeichnet, daß Verzögerungsmittel (16, 27) zur Verzögerung des ersten Sprachsignals ($x_2(i)$) mit einer festen Verzögerungszeit (T_{max}) vorgesehen sind.

8. Mobilfunkendgerät nach einem der Ansprüche 2 bis 7,
dadurch gekennzeichnet,
daß die Sprachverarbeitungsvorrichtung in eine Freisprecheinrichtung (M1, M2, M3, 51, 52, 53) integriert ist.

5

Claims

1. A speech processor provided for processing a first ($x_2(i)$) and at least a further speech signal ($x_1(i)$, $x_3(i)$) consisting of noise and speech signal components and available as sample values, comprising delay means (4, 23, 24) for delaying the sampled further speech signal ($x_1(i)$, $x_3(i)$), comprising control means (3, 26)
- for forming gradient estimates ($\text{grad}(i)$, $\text{sgrad}(i)$) by multiplying error values ($e_{12}(i)$, ($e_{32}(i)$, ($e_{13}(i)$, ($e_{31}(i)$) for two speech signals (for example, $x_1(i)$ and $x_2(i)$) by the output values of a digital filter (6), which filter causes a 90° phase shift to occur and is used for filtering one of the two speech signals (for example, $x_2(i)$),
 - for recursively determining delay estimates ($T_1'(i)$, $T_3'(i)$) from the gradient estimates ($\text{grad}(i)$, $\text{sgrad}(i)$), while the delay values ($T_2(i)$, $T_3(i)$) used for setting the delay means (4, 23, 24) are formed from the delay estimates ($T_1'(i)$, $T_3'(i)$) via a rounding operation, and
 - for forming at least one respective error value ($e_{12}(i)$, ($e_{32}(i)$, ($e_{13}(i)$, ($e_{31}(i)$) for a specific sampling instant (i) from the difference between a speech signal estimate ($x_{1\text{int}}(i)$, $x_{3\text{int}}(i)$) which estimate is used for estimating the further speech signal ($x_1(i)$, $x_3(i)$) at an instant shifted in time by the delay estimate ($T_1'(i)$, $T_3'(i)$) relative to the specific sampling instant (i), and is formed by interpolating sample values of the further speech signal ($x_1(i)$, $x_3(i)$) and the sample value of another one of the speech signals to be processed ($x_1(i)$, $x_2(i)$, $x_3(i)$) at the specific sampling instant (i), and an adder device (5, 25) for adding together the mutually time-shifted speech signals ($x_1(i)$, $x_2(i)$, $x_3(i)$).
2. A mobile radio terminal unit comprising a speech processor as claimed in Claim 1.
3. A mobile radio terminal unit as claimed in Claim 2, characterized in that the digital filter (6) is a digital Hilbert transform.
4. A mobile radio terminal unit as claimed in Claim 3, characterized in that smoothing means (12) are provided for smoothing the gradient estimates ($\text{grad}(i)$).
5. A mobile radio terminal unit as claimed in one of the Claims 2 to 4, characterized in that the speech processor is provided for processing three speech signals ($x_1(i)$, $x_2(i)$, $x_3(i)$).
6. A mobile radio terminal unit as claimed in one of the Claims 2 to 5, characterized in that a linear combination of error values ($e_{12}(i)$ with ($e_{13}(i)$, $e_{31}(i)$ with $e_{32}(i)$) is used for determining a delay estimate ($T_1'(i)$, $T_3'(i)$) for the further speech signal ($x_1(i)$, $x_3(i)$).
7. A mobile radio terminal unit as claimed in one of the Claims 2 to 6, characterized in that the delay means (16, 27) are provided for delaying the first speech signal ($x_2(i)$) by a fixed delay time (T_{max}).
8. A mMobile radio terminal unit as claimed in one of the Claims 2 to 7, characterized in that the speech processor is integrated with a hands-free facility (M1, M2, M3, 51, 52, 53).

Revendications

1. Dispositif de traitement de la parole en vue du traitement d'un premier ($x_2(i)$) et d'au moins un autre ($x_1(i)$, $x_3(i)$) signaux de parole composés de composantes de signal de bruit et de signal de parole et présent sous forme de valeurs de balayage avec des moyens de retardement (4, 23, 24) en vue du retardement du signal de parole supplémentaire balayé ($x_1(i)$, $x_3(i)$) avec des moyens de commande (3, 26):
- en vue de la formation de valeurs estimées de gradient ($\text{grad}(i)$, $\text{sgrad}(i)$) par multiplication de valeurs d'erreur ($e_{12}(i)$, $e_{32}(i)$, $e_{13}(i)$, $e_{31}(i)$) pour deux signaux de parole (par exemple $x_1(i)$ et $x_2(i)$) avec les valeurs de sortie d'un filtre numérique (6) qui provoque un déphasage de 90° et sert au filtrage d'un des deux signaux de parole (par exemple $x_2(i)$)

- en vue de la détermination récurrente de valeur estimées du retardement ($T1'(i)$, $T3'(i)$) à partir des valeurs estimées de gradient ($grad(i)$, $sgrad(i)$), les valeurs de retardement ($T2(i)$, $T3(i)$) qui servent au réglage des moyens de retardement (4, 23, 24) étant formées à partir des valeurs estimées de retardement ($T1'(i)$, $T3'(i)$) par arrondissement et
- en vue de la formation d'au moins une valeur d'erreur ($e_{12}(i)$, $e_{32}(i)$, $e_{13}(i)$, $e_{31}(i)$) pour un moment de balayage déterminé (i) à partir de la différence entre une valeur estimée du signal de parole supplémentaire ($x1_{int}(i)$, $x3_{int}(i)$) qui sert à l'estimation du signal de parole supplémentaire ($x1(i)$, $x3(i)$) à un moment décalé par rapport au moment de balayage(i) déterminé de la valeur estimée du retardement ($T1'(i)$, $T3'(i)$) et est formé par interpolation de valeurs de balayage du signal de parole supplémentaire ($x1(i)$, $x3(i)$) et la valeur de balayage d'un autre des signaux de parole à traiter ($x1(i)$, $x2(i)$, $x3(i)$) au moment du balayage(i) déterminé.

2. Radiotéléphone mobile avec un dispositif de traitement de la parole selon la revendication 1.

3. Radiotéléphone mobile selon la revendication 2,

caractérisé en ce

que le filtre numérique (6) est un transformateur numérique de Hilbert.

4. Radiotéléphone mobile selon la revendication 3,

caractérisé en ce

que les moyens 12 sont prévus en vue du lissage des valeurs estimées de gradient ($grad(i)$).

5. Radiotéléphone mobile selon l'une des revendications 2 à 4,

caractérisé en ce

que le dispositif du traitement de la parole est prévu pour le traitement de trois signaux de parole ($x1(i)$, $x2(i)$, $x3(i)$).

6. Radiotéléphone mobile selon l'une des revendications 2 à 5,

caractérisé en ce

que, pour la détermination d'une valeur estimée du retardement ($T1'(i)$, $T3'(i)$) pour l'autre signal de parole ($x1(i)$, $x3(i)$), l'utilisation d'une combinaison linéaire de valeurs d'erreur ($e_{12}(i)$ avec $e_{13}(i)$ $e_{31}(i)$ avec $e_{32}(i)$) est prévue.

7. Radiotéléphone mobile selon l'une des revendications 2 à 6,

caractérisé en ce

que des moyens de retardement (16, 27) sont prévus en vue du retardement du premier signal de parole ($x2(i)$) avec un temps de retardement fixe (T_{max}).

8. Radiotéléphone mobile selon l'une des revendications 2 à 7,

caractérisé en ce

que le dispositif de traitement de la parole est intégré dans un dispositif à main libre (M1, M2, M3, 51, 52, 53).

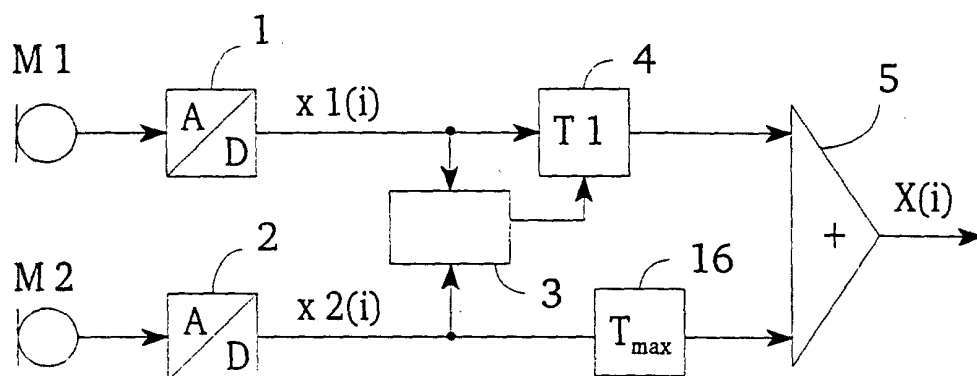


FIG. 1

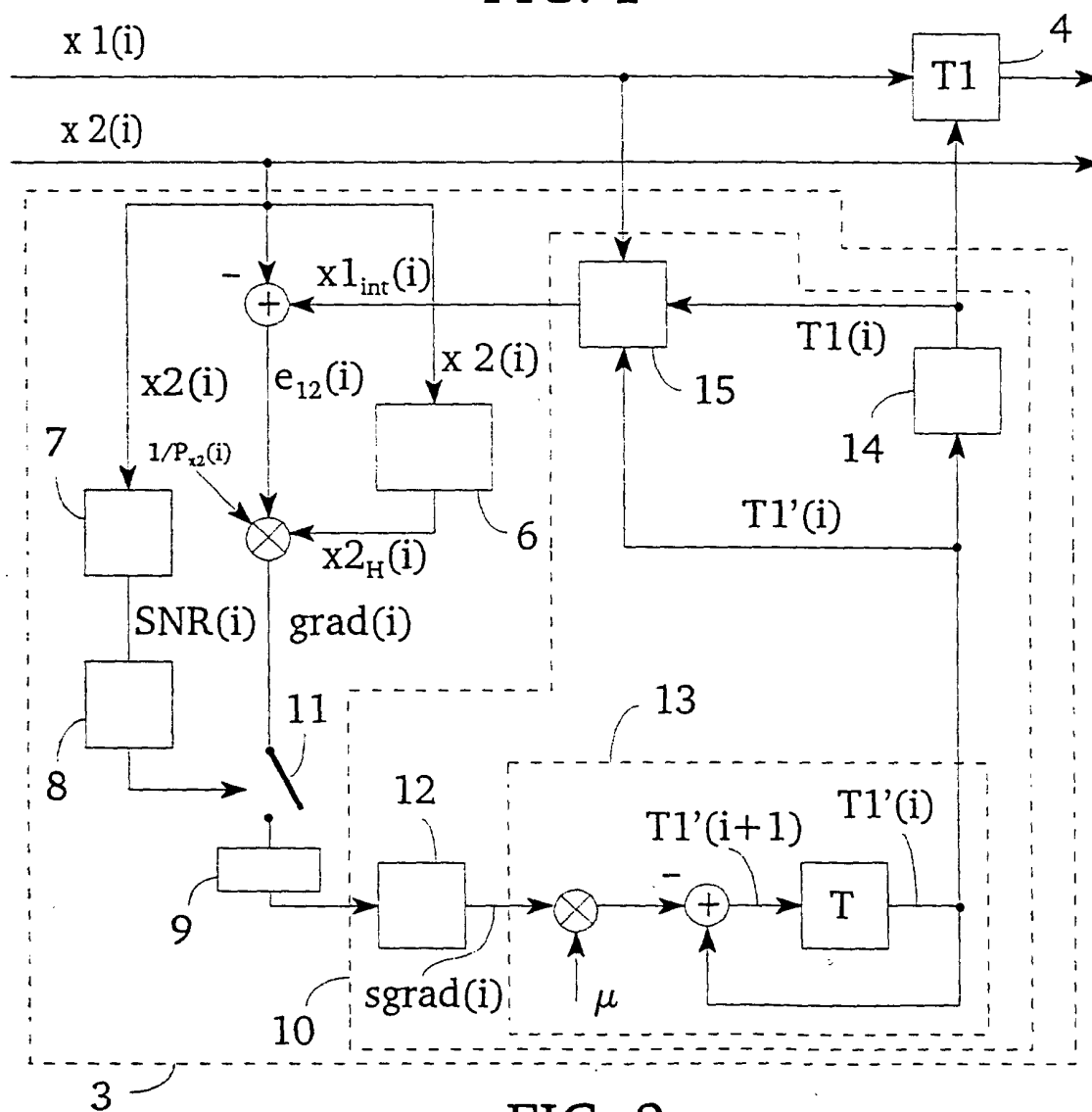


FIG. 2

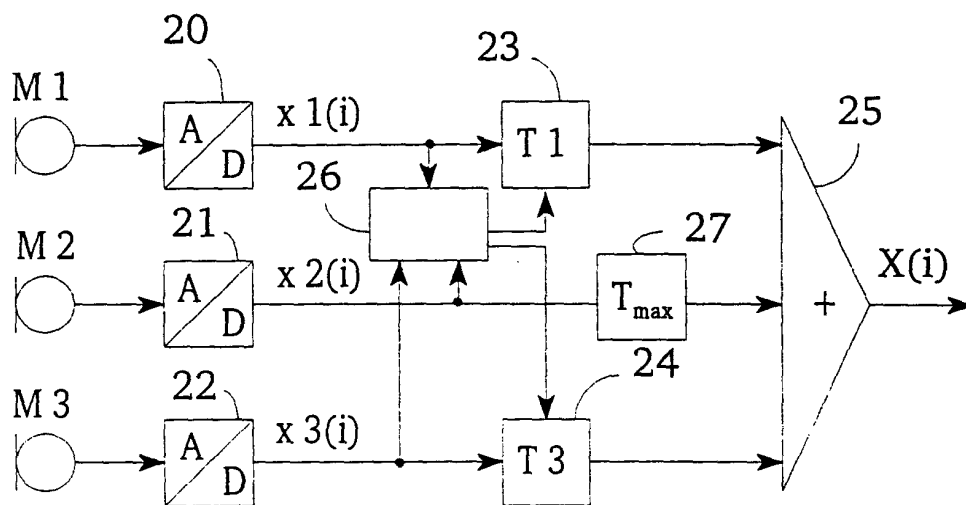


FIG. 3

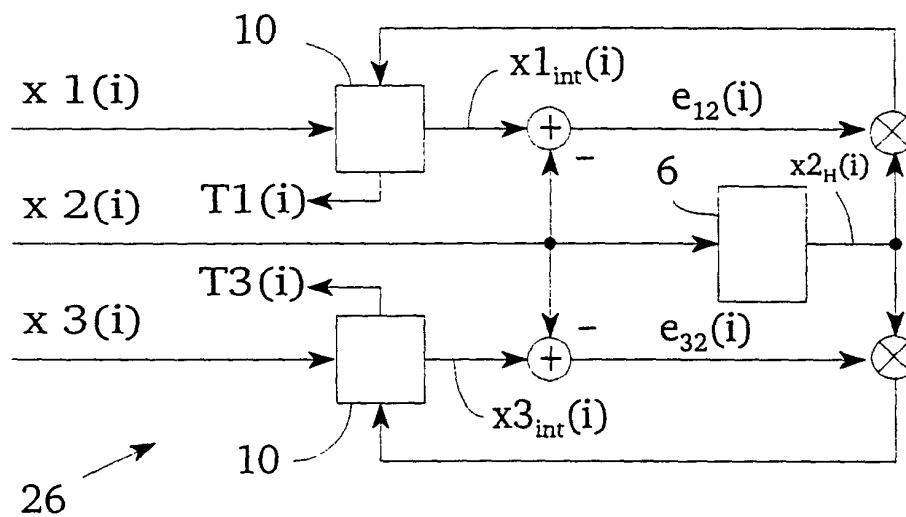


FIG. 4

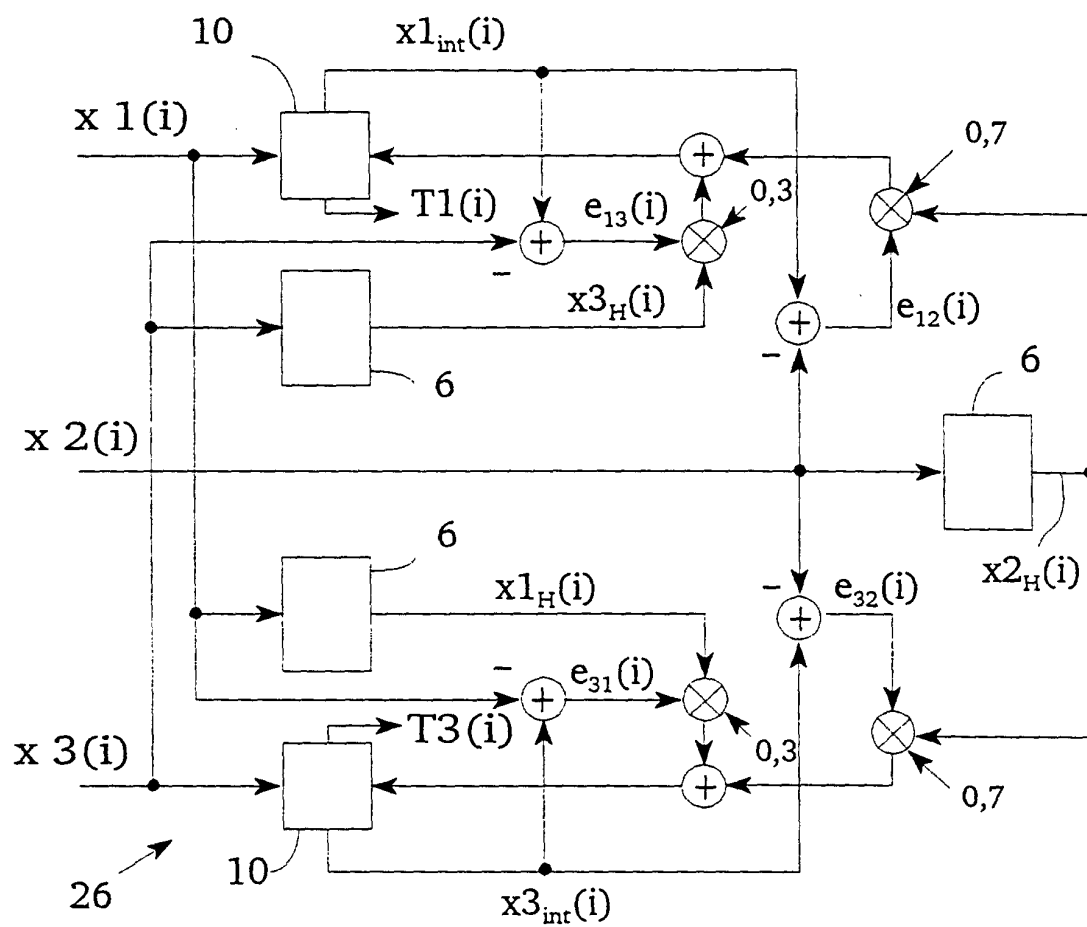


FIG. 5

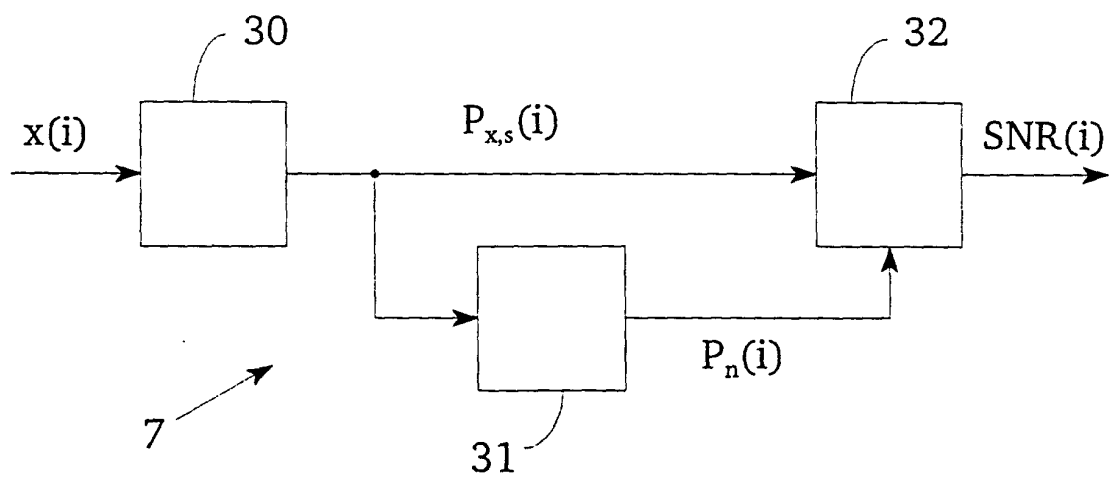


FIG. 6

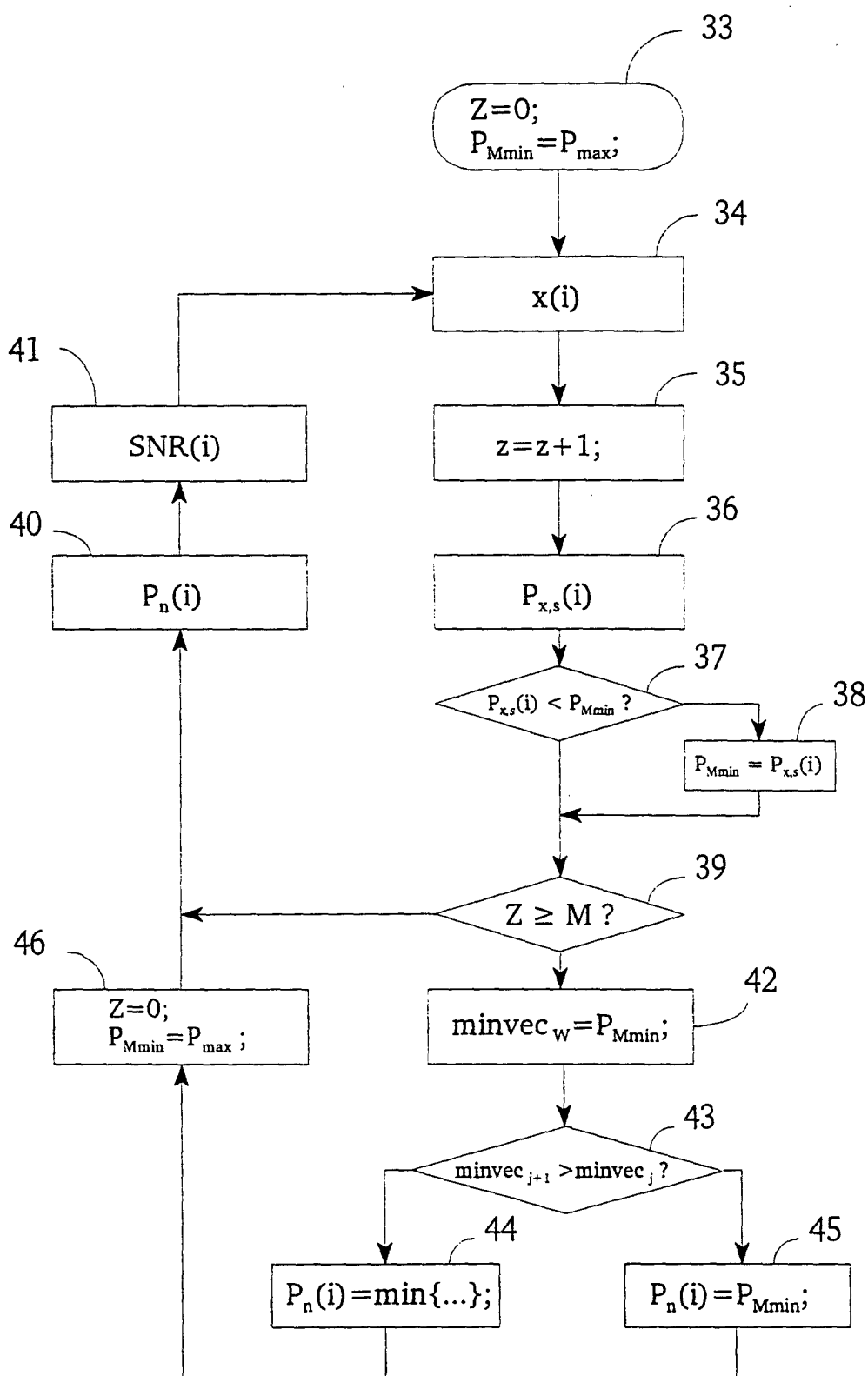


FIG. 7

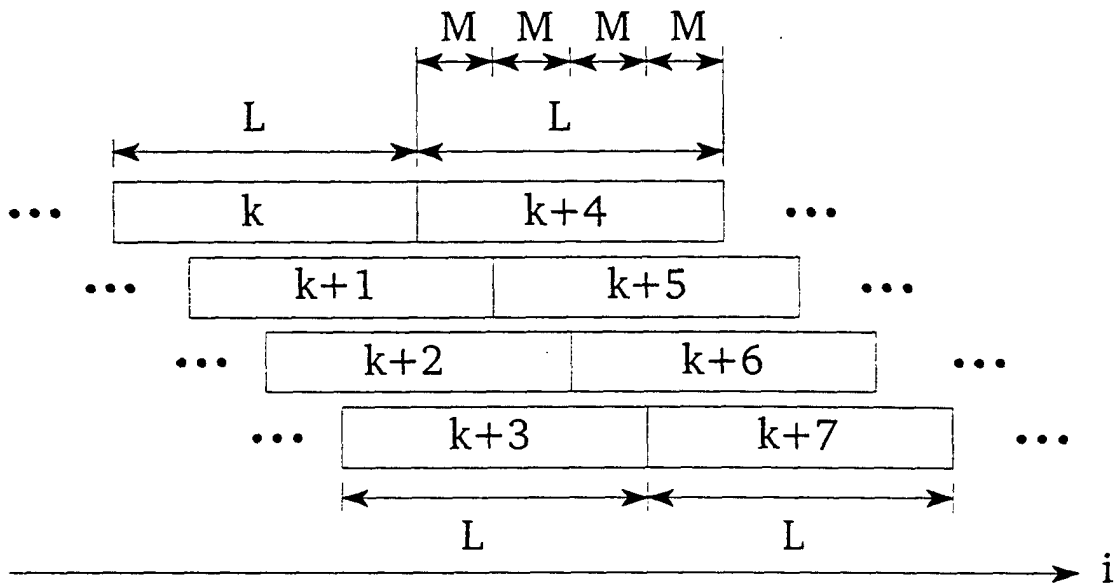


FIG. 8

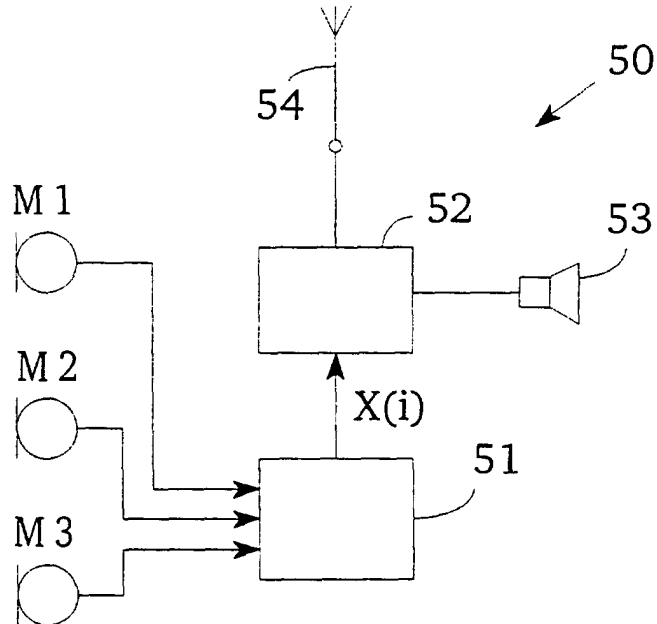


FIG. 9