



(11) **EP 2 080 197 B1**

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
12.01.2011 Patentblatt 2011/02

(51) Int Cl.:
G10L 21/02 (2006.01)

(21) Anmeldenummer: **07800169.0**

(86) Internationale Anmeldenummer:
PCT/AT2007/000424

(22) Anmeldetag: **06.09.2007**

(87) Internationale Veröffentlichungsnummer:
WO 2008/031124 (20.03.2008 Gazette 2008/12)

(54) **VORRICHTUNG ZUR GERÄUSCHUNTERDRÜCKUNG BEI EINEM AUDIOSIGNAL**

APPARATUS FOR NOISE SUPPRESSION IN AN AUDIO SIGNAL

DISPOSITIF D'ÉLIMINATION DU BRUIT DANS UN SIGNAL AUDIO

(84) Benannte Vertragsstaaten:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC MT NL PL PT RO SE SI SK TR

(56) Entgegenhaltungen:
US-A1- 2001 005 822

(30) Priorität: **15.09.2006 AT 15472006**

(43) Veröffentlichungstag der Anmeldung:
22.07.2009 Patentblatt 2009/30

(73) Patentinhaber:
• **Technische Universität Graz**
8010 Graz (AT)
• **Forschungsholding TU Graz GmbH**
8010 Graz (AT)

- **KAWAMURA A ET AL: "A NOISE REDUCTION METHOD BASED ON LINEAR PREDICTION ANALYSIS" ELECTRONICS & COMMUNICATIONS IN JAPAN, PART III - FUNDAMENTAL ELECTRONIC SCIENCE, WILEY, HOBOKEN, NJ, US, Bd. 3, part-3, Nr. 86, 2003, Seiten 1-10, XP001141501 ISSN: 1042-0967 in der Anmeldung erwähnt**
- **MAKHOUL J I & COSELL L K: "Adaptive lattice analysis of speech" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, Bd. CAS-28, Nr. 6, Juni 1981 (1981-06), Seiten 494-499, XP002457971 Bolt Beranek & Newman Inc., Cambridge, MA; USA ISSN: 0098-4094 in der Anmeldung erwähnt**
- **RANKE & KUBING: "Lattice LP Filtering for Noise Reduction in Speech Signals" PROC. INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (INTERSPEECH - ICSLP), 21. September 2006 (2006-09-21), Seiten 265-268, XP001536700**

(72) Erfinder:
• **RANK, Erhard**
1120 Wien (AT)
• **KUBIN, Gernot**
8046 Graz (AT)

(74) Vertreter: **Matschnig, Franz**
Patentanwaltskanzlei Matschnig & Forsthuber OG
Siebensterngasse 54
1071 Wien (AT)

Anmerkung: Innerhalb von neun Monaten nach Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents im Europäischen Patentblatt kann jedermann nach Maßgabe der Ausführungsordnung beim Europäischen Patentamt gegen dieses Patent Einspruch einlegen. Der Einspruch gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist. (Art. 99(1) Europäisches Patentübereinkommen).

EP 2 080 197 B1

Beschreibung

[0001] Die Erfindung bezieht sich auf eine Vorrichtung nach dem Oberbegriff des Hauptanspruches.

[0002] Die Geräuschunterdrückung bei Audiosignalen, insbesondere auch bei Sprachsignalen gewinnt immer mehr an Bedeutung, so in der Telephonie, der automatischen Spracherkennung oder, als nur eines von vielen weiteren Beispielen, bei digitalen Hörgeräten. Geräusche, die es zu unterdrücken gilt, sind in erster Linie nicht-weißes Rauschen, nichtstationäre Geräusche und stark impulsartige Geräusche.

[0003] Verfahren zur Geräuschunterdrückung für Audiosignale sind in einer großen Vielfalt bekannt geworden, und als Beispiele seien genannt: Verfahren, bei welchen leise Audiosignale zunächst angehoben und später wieder abgesenkt werden, wie Pre/Deemphasis für den Rundfunk oder die Dolby-Rauschunterdrückungsverfahren für Tonträgeraufzeichnungen. Weiters Verfahren der spektralen Subtraktion, bei welchen z.B. in Gesprächspausen das Geräusch geschätzt und dann von dem Eingangssignal abgezogen wird. Zu den letztgenannten Verfahren zählen auch Wiener-Filter sowie Ephraim-Malah Filter mit adaptiver Verstärkung für in mehrere Transformationskanäle aufgespaltete Signale. Die bekannten Verfahren sind zum Teil nicht sehr wirkungsvoll, da sie von einem sehr vereinfachten Modell des Geräuschsignals ausgehen, oder sie führen auf Grund einer blockweisen Verarbeitung des Eingangssignals zu Artefakten, die sich als unangenehme Nebengeräusche bemerkbar machen, als so genannte Musical Tones, die im Signal nach der Geräuschreduktion verbleiben. Viele Verfahren führen auch zu einer relativ großen Verzögerung des Ausgangssignals.

[0004] Der Stand der Technik, von welchem die vorliegende Erfindung ausgeht, sind lineare Prädiktionsfilter (LP-Filter), in direkter oder in Lattice-Form (Kreuzglied-Kettenfilter), bei welchen Eigenschaften des gesamten Eingangssignals zur Einstellung der Filterkoeffizienten herangezogen werden. Eine entsprechende Vorrichtung, wie z.B. in der US 2001/0005822 A1 offenbart, besteht aus einem Lattice-Filter, dem ein Eingangssignal $y(n)$ zugeführt ist, welches eine Sprach/Audiokomponente sowie eine Geräuschkomponente aufweist. Für die Einstellung der Komponenten ist eine Koeffizientenberechnungseinheit KBE vorgesehen, welcher die Vorwärts- und Rückwärtsfehlersignale, diese beinhalten in der ersten Stufe auch das Eingangssignal, zugeführt werden. Die Koeffizientenberechnungseinheit liefert dem Filter sodann immer im Sinne einer Minimierung des Prädiktionsfehlers aktualisierte Filterkoeffizienten. Eine Geräuschreduktion unter Verwendung linearer Prädiktionsfilter ist unter anderem auch in der GB 1 520 148 A oder in der US 4, 587, 620 offenbart. Die Verfahren bzw. Vorrichtungen nach dem Stand der Technik gehen immer von dem Eingangssignal aus, ohne dass die besonderen Eigenschaften des Sprachsignals einerseits und des Geräusches andererseits ausreichend Berücksichtigung finden.

[0005] Bei der Vorrichtung nach der US 7, 065,468 wird das Eingangssignal dazu herangezogen, die Koeffizienten für das Prädiktionsfilter zu bestimmen (Einheit 212 bzw. 312 und 318) und aufgrund dieser Koeffizienten erfolgt dann eine Schätzung des Sprachsignals wobei - unter Verwendung einer zusätzlichen Sprachaktivitätsschätzungseinheit (VAD, Einheit 232 bzw. 332) eine Schätzung des Geräuschsignals, nämlich in einer Einheit 234 bzw. 334 dazu erfolgt, mittels eines weiteren zweiten Filters (Einheit 240 bzw. 340) eine Geräuschunterdrückung vorzunehmen.

[0006] Wesentliche Merkmale sind dabei, dass die Schätzung der Koeffizienten des linearen Prädiktionsfilters (Einheit 214) bzw. der linearen Prädiktionsfilter (Einheit 314 und 320 in Fig. 3) nur unter Verwendung des Eingangssignals (bzw. des Fehlersignals des ersten Prädiktionsfilters $e_{ST}(n)$ (Pfad 316)) vorgenommen wird. Ein weiterer wesentlicher Unterschied zur Erfindung liegt darin, dass eine Sprachaktivitätsschätzung unabhängig vom linearen Prädiktionsfilter 214 von den linearen Prädiktionsfiltern 314 und 320 ausgeführt wird, sowie, dass das eigentliche Filter zur Geräuschunterdrückung (Einheit 240 bzw. 340) selbst kein lineares Prädiktionsfilter darstellt. Dabei erfolgt die Geräuschschätzung ("Update Noise Model") nur auf der Basis des Prädiktionsfehlers (vgl. Anspruch 1 der US 7, 065,468), wogegen, wie weiter unten beschrieben, bei der Erfindung die Geräuschschätzung anhand der internen Signale des LP-Fehler-Filters durchgeführt wird. Dieser Unterschied ist auch an Hand der Fig. 2 der US 7, 065,468 ersichtlich.

[0007] Insgesamt zeigt die US 7, 065,468 eine Struktur, die sich bereits von dem Stand der Technik nach der US 2001/0005822 A1 unterscheidet und die noch dazu - verglichen mit der Erfindung - aufwändiger ist. In der US 7, 065,468 wird letztlich ein anderer Weg beschritten, der den Fachmann, dem die US 2001/0005822 A1 bekannt ist und der sich die der Erfindung zugrunde liegende Aufgabe gestellt hat, in eine andere Richtung als jene der Erfindung führen würde.

[0008] Als Publikationen auf diesem Gebiet seien weiters nachstehend genannt

[1] J. D. Markel and A. H. Gray, Jr., Linear Prediction of Speech. Berlin, Heidelberg, New York: Springer, 1976.

[2] J. I. Makhoul and L.K. Cosell, "Adaptive lattice analysis of speech," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, pp.654-659, June 1981.

[3] M. L. Honig and D. G. Messerschmitt, Adaptive Filters: Structures, Algorithms, and Applications. Boston-The Hague-London-Lancaster: Kluwer Academic Publishers, 1984.

[4] A. Kawamura, K. Fujü, Y. Itoh, and Y. Fukui, "A noise reduction method based on linear prediction analysis,"

Electronics and Communications in Japan, Part 3, vol. 86, no. 3, pp. 1-10, 2003.

[5] M. H. Savoji, "Effective noise reduction of speech signals using adaptive lattice filtering, segmentation and soft decision," in IEE Colloquium on New Directions in Adaptive Signal Processing, pp. 7/1-7/5, Feb. 1993.

5 **[0009]** Eine Aufgabe der Erfindung liegt in der Schaffung einer Vorrichtung zur Geräuschunterdrückung für Audiosignale, insbesondere für Sprachsignale, welche ein praktisch unverzögertes Ausgangs-Audiosignal liefert, das auch keine störenden Artefakte mit sich führt.

10 **[0010]** Diese Aufgabe wird mit einer Vorrichtung der eingangs genannten Art gelöst, welche erfindungsgemäß die Merkmale des Kennzeichens des Anspruches 1 aufweist.

[0011] Zweckmäßige Weiterbildungen der Erfindung sind in den abhängigen Ansprüchen 2 bis 9 gekennzeichnet.

[0012] Die Erfindung samt weiteren Vorteilen ist im folgenden an Hand beispielsweise Ausführungsformen näher erläutert, die in der Zeichnung veranschaulicht sind. In dieser zeigen

15 Fig. 1 ein Lattice-Filter nach dem Stand der Technik,

Fig. 2 aus dem geräuschbehafteten Signal ohne Korrektur berechnete Werte des Reflexionskoeffizienten in einem Diagramm,

20 Fig. 3 a und b den Frequenzgang von im Rahmen der Erfindung verwendeten Tiefpässen,

Fig. 4 den zeitlichen Verlauf sowie das Spektrogramm eines geräuschbehafteten Eingangssignals (oben) und dessen Geräuschbefreiung (unten) nach der Erfindung,

25 Fig. 5 das Blockschaltbild einer prinzipiellen Vorrichtung zur Geräuschbefreiung nach dem Stand der Technik

Fig. 6 das Blockschaltbild einer ersten Ausführungsform einer Vorrichtung nach der Erfindung,

30 Fig. 7 das Blockschaltbild einer zweiten Ausführungsform einer Vorrichtung nach der Erfindung und

Fig. 8 das Blockschaltbild einer prinzipiellen Vorrichtung zur Geräuschbefreiung nach dem Stand der Technik unter Verwendung eines LP-Fehler-Filters in direkter Filterform (DFF).

35 **[0013]** Lineare Prädiktion wird üblicherweise auf ein Sprachsignal $x(n)$ angewendet, beispielsweise um die Varianz eines Sprachsignals für dessen Übertragung zu reduzieren. Um einen Signalwert vorher zu sagen, werden langsam zeitveränderliche FIR-Filter (finite impulse response filter, Filter mit endlicher Impulsantwort) niedriger Ordnung verwendet

40

$$\hat{x}(n) = \sum_{i=1}^M b_i(n)x(n-i). \quad (1)$$

45 **[0014]** Hier bedeuten M die Ordnung des LP-Filters und $b_i(n)$ die Filterkoeffizienten, die auf Basis der Signaleigenschaften geschätzt und rahmenweise, z.B. alle 10 ms, aufgefrischt werden. Algorithmen welche die Koeffizienten $b_i(n)$ für das Filter direkt ergeben sind das so genannte "Autokorrelationsverfahren" oder das "Kovarianzverfahren". Die zur Modellierung der spektralen Hüllkurve üblicherweise verwendete Filterordnung liegt, abhängig von der Abtastrate, bei $M=10...20$.

50 **[0015]** Ein dem eben beschriebenen direkten FIR-Filter ebenbürtiges ist das bereits genannte Lattice-Filter, welches einen unmittelbaren Bezug zum menschlichen Stimmtrakt besitzt [1].

[0016] Charakterisiert ist ein solches Filter durch die Gleichungen

55

$$f_0(n) = b_0(n) = x(n),$$

$$f_m(n) = f_{m-1}(n) + k_m(n)b_{m-1}(n-1), \quad (2)$$

5

$$b_m(n) = b_{m-1}(n-1) + k_m(n)f_{m-1}(n). \quad (3)$$

10 welche zu jedem Zeitpunkt n für alle Gliederstufen $m = 1 \dots M$ berechnet werden. Dabei bedeuten $f_m(n)$ und $b_m(n)$ den Vorwärts- bzw. Rückwärts-Fehler in der Stufe m zur Zeit n und $k_m(n)$ die Reflexionskoeffizienten des Filters. Hier sei angemerkt, dass bei einer verallgemeinerten Darstellung die Reflexionskoeffizienten in (2) und (3) unterschiedlich sind, hier jedoch gleiche Reflexionskoeffizienten für vorwärts und rückwärts verwendet werden. In Fig. 1 ist ein Lattice-Filter wie eben beschrieben schematisch dargestellt. Auf ein solches Filter bezieht sich beispielsweise Patentanspruch 6.

15 **[0017]** Der Vorwärtsfehler in der Stufe M ist das Vorhersage-Fehlersignal des LP-Filters:

$$f_M(n) = e(n) = x(n) - \hat{x}(n). \quad (4)$$

20

[0018] Optimale Reflexionskoeffizienten k_m zur Minimierung des quadratischen Mittelwertes des Vorhersagefehlers eines unverzerrten Signals ergeben sich als:

25

$$k_m(n) = -\frac{r_{m-1}(n)}{q_{m-1}(n)}. \quad (5)$$

30

mit den erwarteten Werten für Vorwärts- und Rückwärts- Fehlerkorrelation bzw. Leistung zu:

35

$$r_m(n) = E\{f_m(n)b_m(n-1)\}, \quad (6)$$

40

$$q_m(n) = \frac{1}{2}E\{f_m^2(n) + b_m^2(n-1)\}. \quad (7)$$

[0019] Üblicherweise werden die Erwartungswertoperatoren E in (6) und (7) unter Verwendung Tiefpass-gefilterter Augenblickswerte von $f_m(n)b_m(n-1)$ bzw. $f_m^2(n) + b_m^2(n-1)$ ausgewertet, beispielsweise mit Hilfe einpoliger rekursiver Tiefpassfilter ("lossy integration", verlustbehaftete Integration, siehe weiter unten).

[0020] Somit erfolgt das Adaptieren des Lattice-Filters zur Formung des sich langsam ändernden Eingangs-Sprachsignals durch Berechnung der Gleichungen (6), (7) und (5) für jeden Zeitpunkt n nach dem Filtern - im Gegensatz zur dem rahmenweisen Aktualisieren der Koeffizienten bei der direkten Form des LP-Filters entsprechend Gleichung (1).

50 **[0021]** Was die Geräuschverminderung betrifft, sei angenommen, dass ein beobachtetes Signal $y(n)$ eines additiven linearen Geräuschmodells vorliege:

55

$$y(n) = x(n) + \varepsilon(n), \quad (8)$$

wobei $x(n)$ die Sprachkomponente und $\varepsilon(n)$ eine additive Hintergrund-Geräuschkomponente sei. Gegenstand der Ge-

räuschreduktion ist es, eine gute Schätzung für die SprachsignalKomponente $x(n)$ zu liefern. Für den vorliegenden einkanaligen Fall beruht diese Schätzung ausschließlich auf der Beobachtung des geräuschbehafteten Signals $y(n)$ d.h. es wird keine zusätzliche Information verwendet, wie beispielsweise ein zweites Signal eines Mikrophons, welches lediglich die Hintergrundgeräusche aufnimmt.

5 **[0022]** Eine Verringerung des Hintergrundgeräusches in Sprachsignalen unter Verwendung linearer Prädiktionsfiltrung kann auf der Annahme basieren, dass die SprachsignalKomponente gut vorhersagbar ist, wogegen die Geräuschkomponente diese Eigenschaft nicht besitzt. Somit kann das vorhergesagte Signal $\hat{x}(n)$ als Schätzung für die Sprachkomponente genommen werden. Während bei den Prädiktionsfiltern in der direkten Form in Gleichung (1) der Ausgang unmittelbar das vorhergesagte Signal ist, wird bei den Lattice-Filtern $\hat{x}(n)$ wirksam als die Differenz zwischen dem
10 Eingangssignal und dem Ausgang des Vorwärts-Prädiktionspfades berechnet

$$\hat{x}(n) = y(n) - e(n) \quad (9)$$

15

$\hat{x}(n)$ ist die Schätzung der Sprachkomponente mit $e(n) = f_M(n)$. Vergleiche Gleichung (4).

[0023] Wenn man ein Lattice-LP-Filter für die Geräuschverminderung nach A. Kawamura et al. verwendet, wo ein Filter hoher Ordnung ($N = 256$) benutzt wird, um die spektrale Feinstruktur der Sprache zu modellieren, kann man
20 beobachten, dass $r_m(n)$ eine große Varianz aufgrund der GeräuschsignalKomponente in den höheren Filterstufen aufweist. Es wird vorgeschlagen, die entsprechende Varianz der Reflexionskoeffizienten dadurch zu reduzieren, dass man einen festen (großen) Wert für die Leistungsschätzungen $q_m(n) = G$ verwendet. Die Verringerung des Betrags der Reflexionskoeffizienten oder die Verringerung der Radien der Nullstellen der LP-Filterübertragungsfunktion wurde auch
25 für andere Zwecke vorgeschlagen, z.B. für die Modellierung der spektralen Hüllkurve oder für eine genauere Schätzung der Formanten.

[0024] Eine Verringerung des Betrags der Reflexionskoeffizienten kann durch die Herleitung von Schätzern für r_m und q_m erfolgen, die den mittleren quadratischen Schätzfehler minimieren. Zur Berechnung der Reflexionskoeffizienten für das Lattice-Filter oder der partiellen Korrelationen, welche gleich $-k_m$ sind und auf den Schätzungen von einem Geräuschsignal beruhen, sei für das zusätzliche Geräuschsignal weißes Rauschen mit der Varianz σ_n^2 angenommen,
30 das auch mit $x(n)$ nicht korreliert sein soll. Dies stellt das am wenigsten informierte Modell (maximale Entropie) dar.

[0025] Es ist zwar diese Annahme für das Umgebungsgeräusch im Allgemeinen nicht realistisch, doch beweist sie die Notwendigkeit einer Korrektur der Reflexionskoeffizienten.

[0026] Die Schätzungen für die Korrelation in Gleichung (6) und die Leistung in Gleichung (7) aus der Berechnung der Reflexionskoeffizienten in Gleichung (5) beruhen nun auf dem geräuschbehafteten beobachteten Signal $y(n)$, und man
35 kann die Notwendigkeit eines Korrekturterms, zeigen, um die Schätzungen für die Reflexionskoeffizienten k_m bezogen auf das geräuschfreie Signal $x(n)$ zu gewinnen.

[0027] Im besonderen erhält man für die Schätzung der Reflexionskoeffizienten \hat{k}_m in der ersten Filterstufe $m=1$ den folgenden Erwartungswert für r_0 :

40

$$\begin{aligned} r_0 &= E\{f_0(n) b_0(n-1)\} \\ &= E\{y(n) y(n-1)\} \\ &= E\{(x(n) + \varepsilon(n)) (x(n-1) + \varepsilon(n-1))\} \\ &= E\{x(n) x(n-1)\} \end{aligned} \quad (10)$$

45

50 **[0028]** Für die Fehlervarianz q_0 in der ersten Filterstufe erhält man

55

$$\begin{aligned}
 q_0 &= \frac{1}{2} E\{f_0^2(n) + b_0^2(n-1)\} \\
 &= \frac{1}{2} E\{y^2(n) + y^2(n-1)\} \\
 &= \frac{1}{2} E\{(x(n) + \varepsilon(n))^2 + (x(n-1) + \varepsilon(n-1))^2\} \\
 &= \frac{1}{2} E\{x^2(n) + x^2(n-1)\} + \sigma_n^2
 \end{aligned}
 \tag{11}$$

[0029] Der resultierende Fehler in den Werten der Reflexionsfaktoren ist in Fig. 2 dargestellt. Genauer gesagt sind hier aus dem geräuschbehafteten Signal ohne Korrektur berechnete Werte des Reflexionskoeffizienten k_1 als Funktion eines a priori Signal/Geräusch-Abstandes für unterschiedliche Werte der Autokorrelation $\rho_{xx}(1)$ des ungestörten Signals $x(n)$ veranschaulicht.

[0030] Was die auf das geräuschfreie Signal $x(n)$ bezogenen Reflexionskoeffizienten betrifft, kann die Korrelations-schätzung aus der geräuschbehafteten Beobachtung unverändert verwendet werden, das heißt $\hat{r}_0 = r_0$ wogegen die berechnete Fehlerleistungsschätzung zu korrigieren ist als

$$\hat{q}_0 = q_0 - \sigma_n^2 \tag{12}$$

und der korrigierte Reflexionskoeffizient berechnet wird als

$$\hat{k}_1 = -\frac{\hat{r}_0}{\hat{q}_0} = -\frac{r_0}{q_0 - \sigma_n^2} \tag{13}$$

[0031] Wenn man einführt: $\gamma = \frac{E\{y^2\}}{\sigma_n^2}$ wobei $\gamma - 1 = \frac{E\{y^2\} - \sigma_n^2}{\sigma_n^2}$ der a posteriori bestimmte Signal-

störabstand ist und beachtet wird dass $q_0 = \frac{1}{2} E\{f_0^2(n) + b_0^2(n-1)\}$ kann man diese Gleichung erneut anschreiben als

$$\hat{k}_1 = -\frac{1}{1 - \frac{\sigma_n^2}{q_0}} \frac{r_0}{q_0} = \frac{1}{1 - \frac{1}{\gamma}} k_1 \tag{14}$$

[0032] Das bedeutet eine Skalierung des Reflexionskoeffizienten k_1 , der ursprünglich für das gestörte Signal $y(n)$

berechnet wurde unter Verwendung der Gleichungen (5), (6) und (7) mit einem Faktor $\frac{1}{1 - \frac{1}{\gamma}}$

[0033] Eine Schätzung der Geräuschleistung σ_n^2 kann erfolgen anhand der Leistung des Ausgangssignals $e(n)$ des LP-Fehler-Filters,

$$\sigma_n^2 = E\{e^2(n)\}$$

5 für das Lattice-Filter ist eine möglich Auswertung des Erwartungswerts gegeben durch die Leistungsschätzung in der letzten Stufe des Lattice Filters $q_{M-1}(n)$:

$$10 \quad \sigma_n^2 = q_{M-1}(n),$$

oder, bei Verwendung der Sprachaktivitätsschätzung (siehe unten) anhand der Leistungsschätzung des Gesamteingangssignales bei fehlender Sprachaktivität:

$$15 \quad \sigma_n^2 = q_0(n), \text{ wenn } v \approx 0$$

20 **[0034]** Für die höheren Lattice-Stufen $m = 2, 3, \dots$ kann die Gleichung (14) verallgemeinert werden, wodurch sich eine Korrektur der übrigen Reflexionskoeffizienten k_m ergibt.

25 **[0035]** Dessen ungeachtet kann man aus dem obigen schließen, dass eine Verminderung der Größe der Reflexionskoeffizienten, das heißt eine Verringerung des Verhältnisses zwischen Korrelation und Leistungsschätzung für die Vorhersage eines Signals $x(n)$ bei Beobachtung eines Signal $y(n)$, welches zusätzliches Geräusch enthält, zweckmäßig ist. Das Auffinden der Korrekturgrößen setzt eine verlässliche Schätzung der Geräuschleistung σ_n^2 voraus. Weiters berücksichtigt das Modell bis jetzt nicht irgendwelche Kenntnisse bezüglich der Eigenschaften der Sprache und des zu erwartenden Geräuschsignals.

30 **[0036]** Die Erfindung schafft nun eine Verfahren bzw. eine Vorrichtung, mit welcher man, basierend auf einfachen Annahmen über die Änderung der Korrelation und der Leistung der Sprach- und Geräuschsignale nach der Zeit eine Korrektur der Reflexionsfaktoren erhält. Wie oben ausgeführt, beruhen die Schätzungen der Fehlerkorrelation (6) und der Fehlervarianz (7) üblicherweise auf einer Tiefpassfilterung der Augenblickswerte. Oft wird auch eine einpolige Tiefpassfilterung (verlustbehaftet Integration) verwendet:

$$35 \quad \tilde{r}_m(n) = \lambda_r \tilde{r}_m(n-1) + f_m(n) b_m(n-1), \quad (15)$$

$$40 \quad \tilde{q}_m(n) = \lambda_q \tilde{q}_m(n-1) + \frac{1}{2} (f_m^2(n) + b_m^2(n-1)), \quad (16)$$

45 mit den gleichen Polen bzw. Integrationsfaktoren $\lambda_r = \lambda_q$ sowohl für die Korrelations- als auch die Leistungsschätzung.

[0037] In Einklang mit der Erfindung sind verschiedene Pollagen $\lambda_q \geq \lambda_r$ erlaubt. Die resultierenden Filterfunktionen

$$50 \quad H_r(z) = \frac{1}{1 - \lambda_r z^{-1}}, \quad H_q(z) = \frac{1}{1 - \lambda_q z^{-1}}, \quad (17)$$

55 für $\lambda_r = 0,99608$ und $\lambda_q = 0,99843$ und eine Abtastrate von 16 kHz sind in Fig.3a dargestellt Man kann sehen, dass das Verhältnis von $\tilde{r}_m(n)$ und $\tilde{q}_m(n)$ bei geringeren Frequenzen beeinflusst wird, das heißt für langsam sich ändernde Korrelation und Leistung, wogegen für raschere Änderungen (über ≈ 10 Hz) das Verhältnis verglichen mit den Schätzungen mit $\lambda_r = \lambda_q$ ungeändert bleibt. Unter der Annahme, dass diese Parameter sich für das Sprachsignal rascher ändern

(unter der Annahme beispielsweise einer Phonemrate von 10 pro Sekunde) als für das Geräuschsignal (stationäres oder langsam sich mit der Zeit änderndes Geräusch) wird das sich ergebende Lattice-Prädiktionsfilter die Sprachsignal-komponente gut vorhersagen, wogegen die Geräuschkomponente unterdrückt wird.

[0038] Was impulsartige Geräusche betrifft, kann man vorsehen, das Verhältnis zwischen Korrelation und Leistungsschätzung auch für hohe Frequenzen zu reduzieren, was beispielsweise durch Verwendung eines zweiten Pols in dem Tiefpassfilter für die Korrelation $H_r(z)$ erfolgen kann. Eine entsprechende Übertragungsfunktion ist in Fig.3b dargestellt.

[0039] Im Einzelnen zeigen die Figuren 3a und b die Frequenzgänge eines Tiefpassfilters für eine Fehlerkorrelation $H_r(z)$ (ausgezogene Linien) und die Varianz $H_q(z)$ (punktirierte Linien) für in Fig. 3a zwei einpolige Tiefpassfilter mit $\lambda_r = 0,99608$ und $\lambda_q = 0,99843$ bzw. Figur 3b einen einpoligen Tiefpass für die Leistungsschätzung von $\tilde{q}(n)$ mit $\lambda_q = 0,99843$ und einen zweipoligen Tiefpass für die Korrelationsschätzung $\tilde{r}(n)$ mit $\lambda_{r1} = 0,99608$ und $\lambda_{r2} = 0,9$. Je größer der Abstand zwischen den beiden Übertragungsfunktionen umso stärker ist die Geräuschunterdrückung.

[0040] Um eine gute Reduzierung des Geräusches zu erreichen, kann unter diesen Umständen die Ordnung M des LP-Filters überraschend niedrig gewählt werden, sogar niedriger als die üblicherweise zur Modellierung der spektralen Hüllkurve von Sprachsignalen verwendete Ordnung. Beispielsweise wurde in dem in Fig. 4a gezeigten Beispiel ein Prädiktor mit der Ordnung $M = 10$ für ein Signal mit einer Abtastrate von 16 kHz verwendet. Dieses Beispiel enthält ein mehrmaliges Auftreten von starken, nicht stationären Geräuschbursts, die dank der Erfindung gut beseitigt werden. Das gezeigte Geräusch stammt aus einer Fabrikhallen-Umgebung, das heißt einer äußerst ungünstigen akustischen Umgebung.

[0041] Die Wirksamkeit der Geräuschunterdrückung kann gesteuert werden durch die Einstellung unterschiedlicher Werte für λ_r (bzw. λ_{r1} und λ_{r2}) und λ_q . Wobei diese in Abhängigkeit von der Signalleistung und der Geräuschleistung gewählt werden:

$$(\lambda_r, \lambda_q) = f(\sigma_x^2, \sigma_n^2), \text{ bzw. } (\lambda_{r1}, \lambda_{r2}, \lambda_q) = g(\sigma_x^2, \sigma_n^2). \quad (18)$$

[0042] Weiters ist es zweckmäßig, die Wirksamkeit der Geräuschunterdrückung anhand einer Schätzung der Sprachaktivität zu steuern. Bei der Verwendung eines LP-Fehler-Filters bietet sich (vgl. auch [5]) die Schätzung der wahrscheinlichen Sprachaktivität als reeller Zahl im Wertebereich von 0 bis 1 anhand der Leistungen des Filtereingangssignals und des Filterausgangssignals an:

$$v = \frac{E\{y^2(n)\} - E\{e^2(n)\}}{E\{e^2(n)\}}$$

für ein Lattice Filter ist eine mögliche Auswertung der Erwartungswerte gegeben durch

$$v = \frac{q_0(n) - q_{M-1}(n)}{q_{M-1}(n)}$$

[0043] Zur Steuerung der Geräuschunterdrückung kann dann ein Faktor

$$k_n = 1 - v,$$

für das Ausgangssignal des Lattice Filters $e(n)$, und/oder ein Faktor

$$k_g = v ,$$

5

für das Ausgangssignal verwendet werden.

[0044] Das LP-Fehler-Filter kann als Filter in direkter Filterform (DFF) ausgebildet sein, welches aus dem Eingangssignal ein Prädiktionssignal an seinem Ausgang erzeugt, wobei ein Subtrahierer das Prädiktionssignal vom Eingangssignal abzieht und somit das Ausgangssignal des LP-Fehler-Filters $e(n)$ erzeugt. Die verzögerten Abtastwerte des Eingangssignals (vgl. Gleichung 1) sowie das Ausgangssignal des Subtrahierers $e(n)$ entsprechen den internen Signalen des LP-Fehler-Filters.

[0045] Ein wichtiges Merkmal der erfindungsgemäßen Geräuschunterdrückung besteht in der an die Eigenschaften des Sprach- und des Geräuschsignals angepassten Auswertung der Erwartungswertoperatoren und damit der optimalen Einstellung der Filterkoeffizienten für das lineare Prädiktionsfilter, sowie in der Sprachaktivitätsschätzung und deren Verwendung bei der Schätzung des Geräuschsignals, und zur Steuerung der Wirksamkeit der Geräuschunterdrückung und der Amplitude des Ausgangssignales.

[0046] Wenngleich es klar sein sollte, dass der Rechenaufwand mit der gewählten Filterordnung steigt und daher der Rechenaufwand unter Umständen größer sein kann, als jener bei Verwendung einer schnellen Fouriertransformation, liegt ein wesentlicher Vorteil der Erfindung darin, dass sie eine Geräuschreduktion ohne Verzögerung des Sprachsignals ermöglicht, was ein besonderer Vorteil ist, vor allem beim Einsatz in Hörgeräten.

20

Patentansprüche

25 1. Vorrichtung zur Geräuschunterdrückung mit einer linearen Prädiktions-Analysierschaltung mit einem LP-Fehler-Filter (LFF), das aus einem Lattice-Filter besteht und auf Basis eines ersten, von Geräusch überlagerten Sprachsignals $y(n) = x(n) + \varepsilon(n)$, wobei $x(n)$ ein nicht von Geräusch überlagertes Sprachsignal und $\varepsilon(n)$ das Geräusch darstellen, ein LP-Fehler-Filter Ausgangssignal $e(n)$ erzeugt, einer Koeffizientenberechnungseinheit (KBE), welche die Koeffizienten des LP-Fehler-Filters auf Basis der Vorwärts und Rückwärtsteilersignale $f_m(n)$ und $b_m(n)$ der Lattice-Filterstufen des LP-Fehler-Filters sowie des Ein- und Ausgangssignals $y(n)$ und $e(n)$ des LP-Fehler-Filters
30 aktualisiert, und einer Subtrahiereinheit, welche das LP-Fehler-Filter-Ausgangssignal $e(n)$ von dem ersten Sprachsignal $y(n)$ in einem Subtrahierer subtrahiert und nach der Subtraktion den Rest als zweites Sprachsignal $\hat{x}(n) = y(n) - e(n)$ ausgibt, in welchem das Geräusch unterdrückt ist, eine Geräuschschätzungseinheit (GSE) vorgesehen ist, welche auf Basis der Signale $f_m(n)$, $b_m(n)$, $y(n)$ und $e(n)$ des LP-Fehler-Filters ein Geräuschleistungssignal σ_n^2 und ein Sprachleistungssignal σ_x^2 erzeugen, die der Koeffizientenberechnungseinheit (KBE) zugeführt sind, in welcher die Koeffizienten des LP-Fehler-Filters zu jedem Zeitpunkt n berechnet werden,

35

dadurch gekennzeichnet, dass

die Koeffizientenberechnungseinheit (KBE) weiters dazu eingerichtet ist, die korrigierte Fehlervarianz \hat{q}_0 gemäß

40

$$\hat{q}_0 = q_0 - \sigma_n^2$$

45 und den korrigierten Reflexionskoeffizienten \hat{k}_1 gemäß

45

$$\hat{k}_1 = -\frac{\hat{r}_0}{\hat{q}_0} = -\frac{r_0}{q_0 - \sigma_n^2}$$

50

zu ermitteln, mit q_0 als Fehlervarianz und r_0 als Reflexionskoeffizient, und die Koeffizientenberechnungseinheit (KBE) weiters dazu eingerichtet ist, die Fehlerkorrelation gemäß

55

$$\tilde{r}_m(n) = \lambda_r \tilde{r}_m(n-1) + f_m(n) b_m(n-1)$$

5

und die Fehlervarianz gemäß

10

$$\tilde{q}_m(n) = \lambda_q \tilde{q}_m(n-1) + \frac{1}{2}(f_m^2(n) + b_m^2(n-1))$$

mit $\lambda_q > \lambda$,

zu ermitteln, wobei λ , und λ_q Pole der für $\tilde{r}_m(n)$ und $\tilde{q}_m(n)$ angegebenen Tiefpassfiltergleichungen sind. .

15

2. Vorrichtung nach Anspruch 1, **dadurch gekennzeichnet, dass** in der Koeffizientenberechnungseinheit (KBE) für die Korrelationsschätzung der Fehlerkorrelation $\tilde{r}(n)$ ein zweipoliger Tiefpass vorgesehen ist.

20

3. Vorrichtung nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** eine Sprachaktivitätsschätzungseinheit (SAE) vorgesehen ist, welche auf Basis der Signale $f_m(n)$, $b_m(n)$, $y(n)$ und $e(n)$ ein Sprachaktivitätssignal v erzeugt, das der Koeffizientenberechnungseinheit (KBE) zugeführt ist und von dieser im Sinne einer Optimierung der Geräuschunterdrückung berücksichtigt wird.

25

4. Vorrichtung nach Anspruch 3, **dadurch gekennzeichnet, dass** die Sprachaktivitätsschätzungseinheit (SAE) zur Bildung eines Geräuschunterdrückungsfaktors (k_n) eingerichtet ist, welcher einem Eingang eines ersten Multiplizierers (MU1) zugeführt ist und dessen anderem Eingang das Ausgangssignal des LP-Fehler-Filters zugeführt ist und der vor der Subtrahiereinheit liegt.

30

5. Vorrichtung nach Anspruch 4, **dadurch gekennzeichnet, dass** die Sprachaktivitätsschätzungseinheit (SAE) zur Bildung eines Gesamtsignalfaktors (k_g) eingerichtet ist, welcher einem Eingang eines zweiten Multiplizierers (MU2) zugeführt ist, dessen anderem Eingang das Ausgangssignal der Subtrahiereinheit zugeführt ist.

35

6. Kaskade, bestehend aus zumindest zwei hintereinander geschalteten Vorrichtungen nach einem der Ansprüche 1 bis 5.

Claims

40

1. Device for noise reduction with a linear predictive analyser circuit with an LP error filter (LFF), which consists of a lattice filter and generates an LP error filter output signal $e(n)$, on the basis of a first speech signal $y(n) = x(n) + \varepsilon(n)$ overwritten with noise, whereby $x(n)$ represents a speech signal not overwritten with noise and $\varepsilon(n)$ represents the noise, a coefficient calculation unit (KBE), which updates the coefficients of the LP error filter on the basis of the input and output signals $y(n)$ and $e(n)$, forward and backward error signals $f_m(n)$ and $b(n)$ of the lattice filter stages of the LP error filter, as well as the LP error filter, and a subtraction unit, which subtracts the LP error filter output signal $e(n)$ from the first speech signal $y(n)$ in a subtractor, and after the subtraction emits the remainder as a second speech signal $\hat{x}(n) = y(n) - e(n)$ in which the noise is reduced, a noise estimation unit (GSE) is provided, which, on the basis of signals $f_m(n)$, $b(n)$, $y(n)$ and $e(n)$ of the LP error filter, generates a noise performance signal σ_n^2 and a speech performance signal σ_x^2 which are fed into the coefficient calculation unit (KBE), in which the coefficients of the LP error filter are calculated at each point n ,

45

characterised in that,

50

the coefficient calculation unit (KBE) is also equipped to calculate the corrected error variance \hat{q}_0 according to

55

$$\hat{q}_0 = q_0 - \sigma_n^2$$

and the corrected reflection coefficients \hat{k} according to

$$\hat{k}_1 = -\frac{\hat{r}_0}{\hat{q}_0} = -\frac{r_0}{q_0 - \sigma_n^2}$$

with q_0 denoting the error variance and r_0 denoting the reflection coefficient,
and the coefficient calculation unit (KBE) is also equipped to calculate the error correlation according to

$$\tilde{r}_m(n) = \lambda_r \tilde{r}_m(n-1) + f_m(n) b_m(n-1)$$

and the error variance according to

$$\tilde{q}_m(n) = \lambda_q \tilde{q}_m(n-1) + \frac{1}{2} (f_m^2(n) + b_m^2(n-1))$$

with $\lambda_r > \lambda_q$

whereby λ_r and λ_q are poles which are for low-pass filter equations denoted as $\tilde{r}_m(n)$ and $q_{\tilde{m}(n)}$.

2. Device according to claim 1, **characterised in that** a bipolar low-pass is provided in the coefficient calculation unit (KBE) for the correlation estimate of the error correlation $r_f(n)$.
3. Device according to claim 1 or 2, **characterised in that** a voice activity estimation unit (SAE) is provided, which on the basis of signals $f_m(n)$ and $b(n)$, $y(n)$ and $e(n)$ generates a voice activity signal v , which is fed into the coefficient calculation unit (KBE) and is considered by the same in terms of optimising noise reduction.
4. Device according to claim 3, **characterised in that** the voice activity estimation unit (SAE) is equipped to generate a noise reduction factor (k_n), which is fed into an input of a first multiplier (MU1), and with the output signal of the LP error filter being fed into the other input thereof, and which lies before the subtraction unit.
5. Device according to claim 4, **characterised in that** the voice activity estimation unit (SAE) is equipped to generate a total signal factor (K_g) which is fed into an input of a second multiplier (MU2), with the output signal of the subtraction unit being fed into the other input thereof.
6. Cascade, consisting of at least two devices connected in series according to any one of claims 1 to 5.

Revendications

1. Dispositif d'atténuation du bruit comprenant un circuit de prédiction et d'analyse linéaire avec un filtre d'erreur passe-bas (LFF) qui se compose d'un filtre lattice et qui, sur la base d'un premier signal vocal $y(n) - x(n) + \varepsilon(n)$ auquel est superposé du bruit, $x(n)$ représentant un signal vocal non superposé de bruit et $\varepsilon(n)$ le bruit, génère un signal de sortie de filtre d'erreur passe-bas $e(n)$, avec une unité de calcul de coefficient (KBE) qui actualise les coefficients du filtre d'erreur passe-bas sur la base des signaux d'erreur avant et arrière $f_m(n)$ et $b_m(n)$ des étages de filtre lattice du filtre d'erreur passe-bas et sur la base du signal d'entrée et du signal de sortie $y(n)$ et $e(n)$ du filtre d'erreur passe-bas, et avec une unité de soustraction qui soustrait le signal de sortie de filtre d'erreur passe-bas $e(n)$ du premier signal vocal $y(n)$ dans un soustracteur et, après soustraction, sort le reste sous la forme d'un deuxième signal vocal $\hat{x}(n) = y(n) - e(n)$ où le bruit est atténué, où une unité d'évaluation du bruit (GSE) est prévue, laquelle génère sur la base des signaux $f_m(n)$, $b_m(n)$, $y(n)$ et $e(n)$ du filtre d'erreur passe-bas un signal de puissance du bruit σ_n^2 et un signal de puissance vocale σ_x^2 , lesdits signaux étant acheminés vers l'unité de calcul de coefficient (KBE) où les coefficients du filtre d'erreur passe-bas sont calculés à chaque moment n ,
caractérisé

EP 2 080 197 B1

en ce que l'unité de calcul de coefficient (KBE) est en outre prévue pour déterminer la variance d'erreur corrigée \hat{q}_0 selon

5

$$\hat{q}_0 = q_0 - \sigma_n^2$$

et le coefficient de réflexion corrigé \hat{k} selon

10

$$\hat{k}_1 = -\frac{\hat{r}_0}{\hat{q}_0} = -\frac{r_0}{q_0 - \sigma_n^2}$$

15

avec q_0 comme variance d'erreur et r_0 comme coefficient de réflexion, et **en ce que** l'unité de calcul de coefficient (KBE) est en outre prévue pour déterminer la corrélation d'erreur selon

20

$$\tilde{r}_m(n) = \lambda_r \tilde{r}_m(n-1) + f_m(n) b_m(n-1)$$

et la variance d'erreur selon

25

$$\tilde{q}_m(n) = \lambda_q \tilde{q}_m(n-1) + \frac{1}{2} (f_m^2(n) + b_m^2(n-1))$$

avec $\lambda_q > \lambda_r$,

λ_r et λ_q étant des pôles des équations de filtre passe-bas données pour $\tilde{r}_m(n)$ et $\tilde{q}_m(n)$.

30

2. Dispositif selon la revendication 1, **caractérisé en ce qu'**un filtre passe-bas bipolaire est prévu dans l'unité de calcul de coefficient (KBE) pour l'évaluation de corrélation de la corrélation d'erreur $\tilde{r}(n)$.
3. Dispositif selon la revendication 1 ou la revendication 2, **caractérisé en ce qu'**une unité d'évaluation d'activité vocale (SAE) est prévue, laquelle génère sur la base des signaux $fm(n)$, $bm(n)$, $y(n)$ et $e(n)$ du filtre d'erreur passe-bas un signal d'activité vocale v qui est acheminé vers l'unité de calcul de coefficient (KBE) et qui est pris en compte par celle-ci en vue d'une optimisation de l'atténuation du bruit.
4. Dispositif selon la revendication 3, **caractérisé en ce que** l'unité d'évaluation d'activité vocale (SAE) est prévue pour former un facteur d'atténuation du bruit (k_n) qui est acheminé vers une entrée d'un premier multiplicateur (MU1), vers l'autre entrée duquel est acheminé le signal de sortie du filtre d'erreur passe-bas, ledit multiplicateur étant situé en amont de l'unité de soustraction.
5. Dispositif selon la revendication 4, **caractérisé en ce que** l'unité d'évaluation d'activité vocale (SAE) forme un facteur de signal global (k_g) qui est acheminé vers une entrée d'un deuxième multiplicateur (MU2), vers l'autre entrée duquel est acheminé le signal de sortie de l'unité de soustraction.
6. Cascade composée d'au moins deux dispositifs selon l'une des revendications 1 à 5, montés l'un derrière l'autre.

50

55

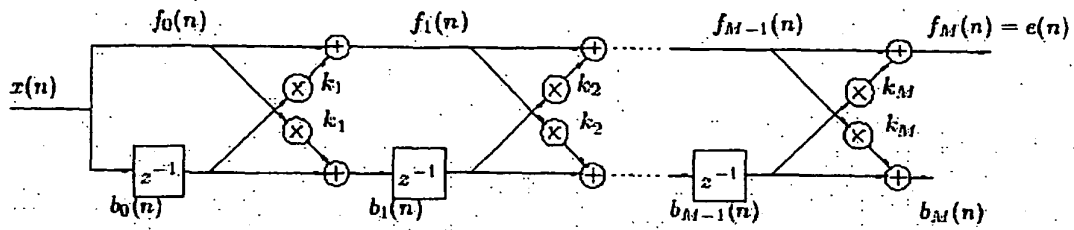


Fig. 1

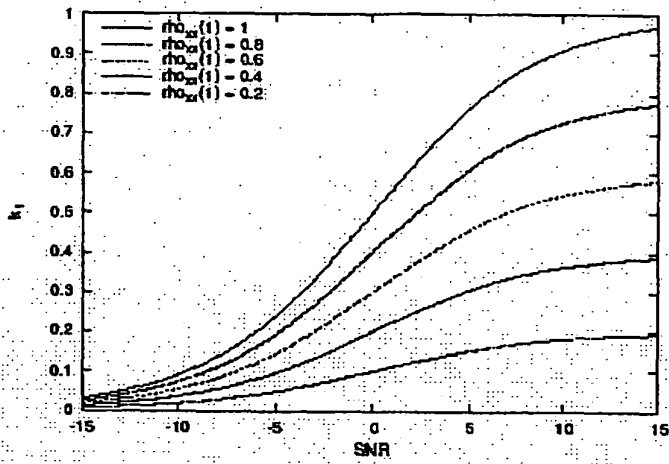


Fig. 2

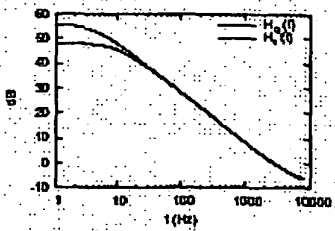


Fig. 3a

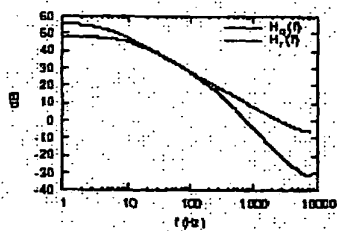


Fig. 3b

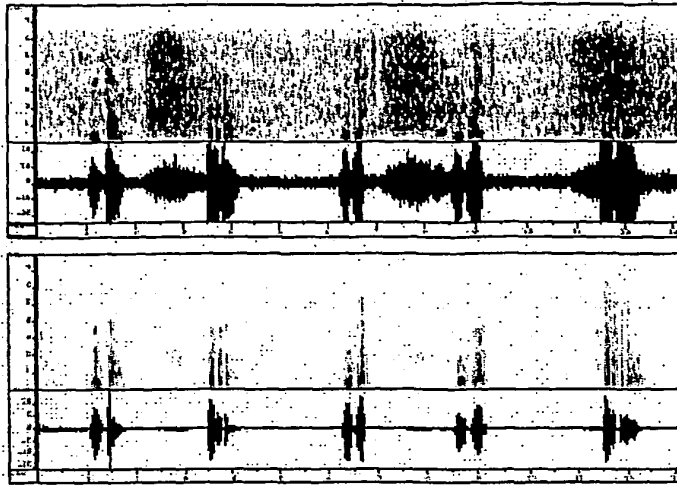


Fig. 4

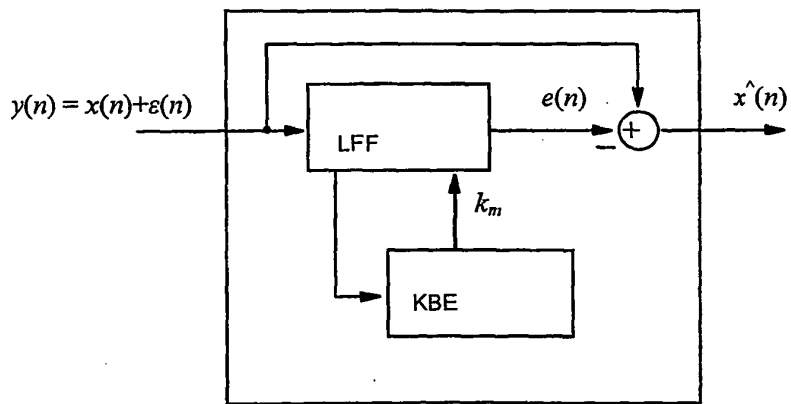


Fig. 5 (Stand der Technik)

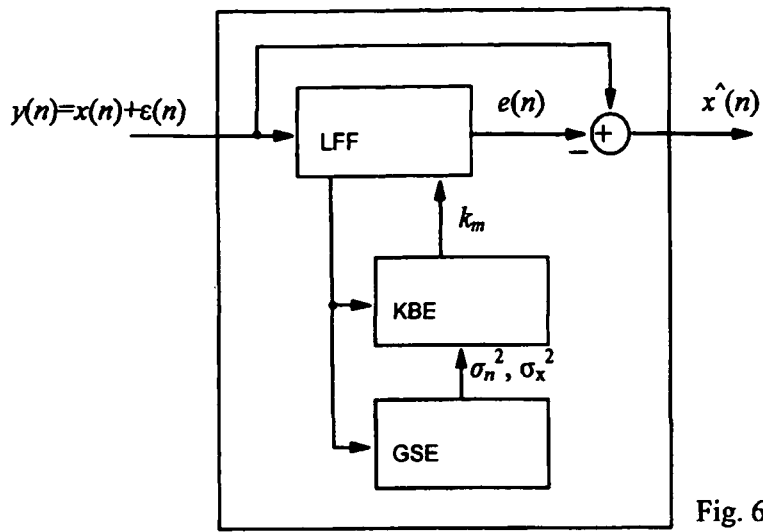


Fig. 6

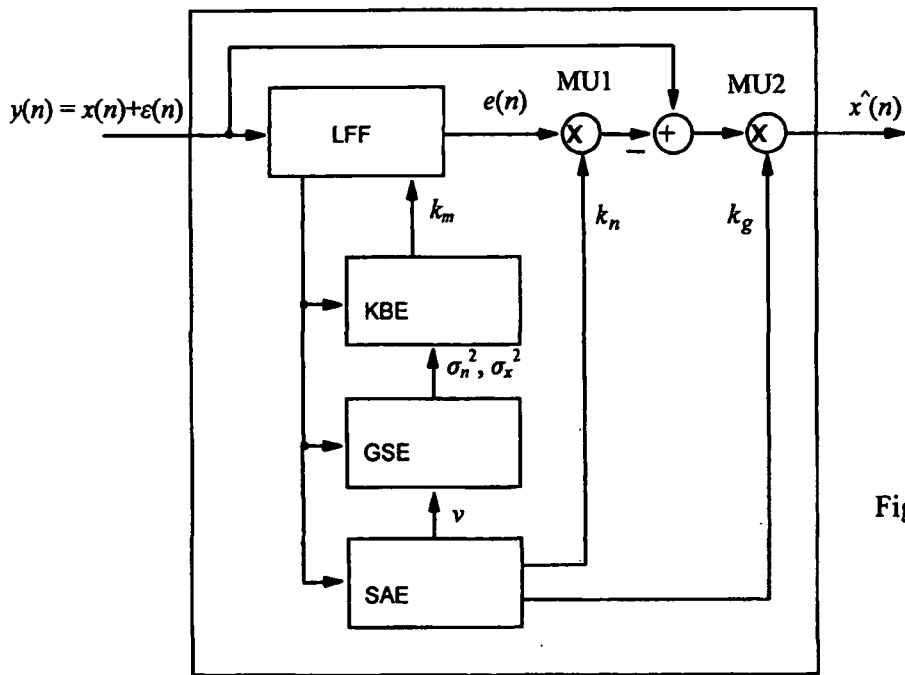


Fig. 7

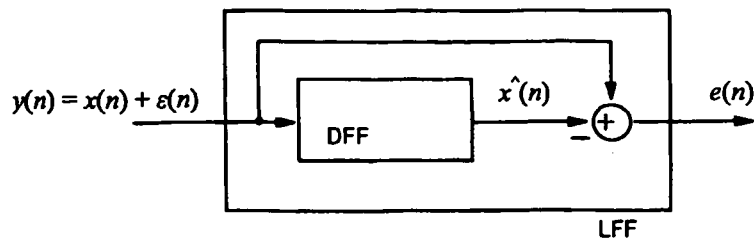


Fig. 8

IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

Diese Liste der vom Anmelder aufgeführten Dokumente wurde ausschließlich zur Information des Lesers aufgenommen und ist nicht Bestandteil des europäischen Patentdokumentes. Sie wurde mit größter Sorgfalt zusammengestellt; das EPA übernimmt jedoch keinerlei Haftung für etwaige Fehler oder Auslassungen.

In der Beschreibung aufgeführte Patentdokumente

- US 20010005822 A1 [0004] [0007]
- GB 1520148 A [0004]
- US 4587620 A [0004]
- US 7065468 B [0005] [0006] [0007]

In der Beschreibung aufgeführte Nicht-Patentliteratur

- **J. D. Markel ; A. H. Gray ; Jr., Linear.** Prediction of Speech. Springer, 1976 [0008]
- **J. I. Makhoul ; L.K. Cosell.** Adaptive lattice analysis of speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1981, vol. 29, 654-659 [0008]
- Boston-The Hague-London-Lancaster. **M. L. Honig ; D. G. Messerschmitt.** Adaptive Filters: Structures, Algorithms, and Applications. Kluwer Academic Publishers, 1984 [0008]
- **A. Kawamura ; K. Fujü ; Y. Itoh ; Y. Fukui.** A noise reduction method based on linear prediction analysis. *Electronics and Communications in Japan*, 2003, vol. 86 (3), 1-10 [0008]
- **M. H. Savoji.** Effective noise reduction of speech signals using adaptive lattice filtering, segmentation and soft decision. *IEE Colloquium on New Directions in Adaptive Signal Processing*, Februar 1993, 7, 1-75 [0008]