



(12) **EUROPÄISCHE PATENTANMELDUNG**

(21) Anmeldenummer : **94102963.9**

(51) Int. Cl.<sup>5</sup> : **G10L 3/02**

(22) Anmeldetag : **28.02.94**

(30) Priorität : **11.03.93 DE 4307688**

(43) Veröffentlichungstag der Anmeldung :  
**14.09.94 Patentblatt 94/37**

(84) Benannte Vertragsstaaten :  
**DE FR GB IT**

(71) Anmelder : **Daimler-Benz Aktiengesellschaft**  
**Postfach 80 02 30**  
**D-70546 Stuttgart (DE)**

(72) Erfinder : **Linhard, Klaus, Dr. Ing.**  
**Kasernstrasse 43**  
**D-89231 Neu-Ulm (DE)**

(74) Vertreter : **Amersbach, Werner, Dipl.-Ing.**  
**AEG Aktiengesellschaft**  
**Postfach 70 02 20**  
**D-60591 Frankfurt (DE)**

(54) **Verfahren zur Geräuschreduktion für gestörte Sprachkanäle.**

(57) Die Erfindung betrifft ein Verfahren, daß nicht nur zur Störfreiung z.B. bei der automatischen Spracherkennung eingesetzt werden kann, sondern auch zur Verbesserung der Sprachqualität für den Menschen z.B. Freisprechen beim Autotelefon. Die Geräuschreduktion wird 2kanalig oder mehrkanalig durchgeführt, in der Art, daß schrittweise, systematisch die zeitlichen und die raumakustischen Signaleigenschaften von Sprache und Störung ausgenutzt werden.

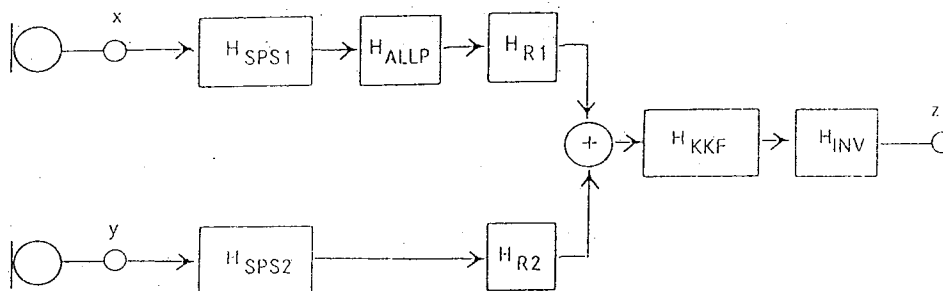


FIG.1

Die Erfindung betrifft ein Verfahren nach dem Oberbegriff des Patentanspruchs 1.

Ein derartiges Verfahren findet Anwendung bei der automatischen Spracherkennung oder bei Freisprechanlagen zur Verbesserung der Sprachqualität, z.B. in Büroräumen oder im Kraftfahrzeug.

Gestörte Sprache ist besser erfaßbar, wenn sie mit zwei oder mehreren Kanälen aufgezeichnet wird. Dabei soll in jedem Kanal Sprache und Störung vorhanden sein. Die mehrkanaligen Signale werden mit einer digitalen Signalverarbeitung aufbereitet.

Bei mehrkanaligen Systemen ist zunächst der Laufzeitunterschied des Nutzsignals in den einzelnen Kanälen zu ermitteln. Dabei wird es später möglich, die einzelnen Kanäle wieder phasenrichtig zu einem Kanal zusammenzuführen.

Von besonderem Interesse sind Systeme mit 2 Kanälen, da sich hiermit bereits ein räumliches Schallfeld nach einzelnen Richtungen auflösen läßt, der Rechenaufwand aber noch erträglich bleibt.

Ist die Richtung bekannt, aus der das interessierende Schallereignis eintrifft, wird eine akustische Richtkeule auf dieses Ereignis eingestellt.

Die Geräuschreduktion wird zunächst in jedem einzelnen Kanal durchgeführt. Da die Geräuschreduktion nicht fehlerfrei arbeitet können Verzerrungen und künstliche Einfügungen (z.B. "musical tones") entstehen. Bei der Zusammenführung der einzelnen verarbeiteten Kanäle ergibt sich eine Mittelung und damit Verringerung dieser Fehler.

Das Summensignal wird anschließend nachverarbeitet, indem die Kreuzkorrelation der Signale in den einzelnen Kanälen verwendet wird. Dabei wird vorausgesetzt daß Störungen oder Nachhall weniger korreliert ist als das Nutzsignal der Kanäle.

Ein Verfahren zur Zusammenführung von 2 gestörten Sprachkanälen ist aus der Veröffentlichung "Multimicrophone signal-processing technique to remove room reverberation from speech signals" von Allen, Berkley und Blauert (J: Acoust. Soc. Am., Vol.62, No. 4, October 1977) und aus "Noise Suppression Signal Processing Using 2-Point Received Signals" von Kaneda und Tohyame (Electronics and Communication in Japan, Vol. 67-A, No. 12, 1984) bekannt. Das erste Verfahren ist zur Enthaltung von Sprachsignalen gedacht und verwendet keinen echten Phasenausgleich des Nutzsignals und die Enthaltung mit Geräuschreduktion wird nur in einer Nachverarbeitungsstufe durchgeführt. Das zweite Verfahren benutzt einen einfachen linearen Phasenausgleich der Kanäle, die Geräuschreduktion erfolgt aber auch hier nur in der Nachverarbeitungsstufe.

Der Erfindung liegt deshalb die Aufgabe zugrunde, ein Verfahren zur Geräuschreduktion anzugeben, bei dem die Geräuschreduktion in mehreren Stufen durchgeführt und eine deutliche Verbesserung der Sprachqualität erzielt wird.

Die Aufgabe wird gelöst durch die im kennzeichnenden Teil des Patentanspruchs 1 angegebenen Merkmale. Vorteilhafte Ausgestaltungen und/oder Weiterbildungen sind den Unteransprüchen zu entnehmen.

Mit dem erfindungsgemäßen Verfahren werden die räumlichen und die zeitlichen Eigenschaften des Nutzsignals und der Störung systematisch ausgenutzt:

1.) räumliche Eigenschaft der Schallfelder:

a) Dämpfung von Punktstörquellen

Mit digitalen Richtungsfiltern am Eingang der Kanäle wird zusammen mit der Phasenschätzung eine akustische Richtkeule auf den Sprecher ausgerichtet. Für die Phasenschätzung wird das in der unveröffentlichten deutschen Patentanmeldung P 42 43 831 beschriebene Verfahren verwendet. Es ist robust gegenüber Störungen und benötigt nur einen geringen Rechenaufwand. Die Richtungsfilter sind fest eingestellt. Es wird angenommen, daß der Sprecher sich relativ nahe an den Mikrofonen befindet (Abstand  $\leq 1\text{m}$ ) und sich nur in einem beschränkten Bereich bewegt. Instationäre und stationäre Punktstörquellen werden durch diese räumliche Auswertung gedämpft.

b) Dämpfung von diffusen Störquellen

In der Nachverarbeitung werden mit Hilfe der Kreuzkorrelation die diffusen Stör- und Hallanteile gedämpft.

2.) zeitliche Signaleigenschaften:

Die spektrale Subtraktion schätzt die Störung in den Sprachpausen und führt eine betragsmäßige Subtraktion im Spektralbereich durch. Hier werden die zeitlich stationären Störanteile gedämpft.

3.) Mittelung der Kanäle (Addition):

Durch die räumliche Trennung der Aufnahmekanäle (Mikrofone in einem bestimmten Abstand) treten Fehler der spektralen Subtraktion (Verzerrung und "musical tones") in den einzelnen Kanälen z.T. zeitlich zufällig auf. Eine Mittelung der Kanäle vermindert diesen Fehler.

Die Erfindung wird anhand von Ausführungsbeispielen näher erläutert und Bezugnahme auf schematische Zeichnungen.

FIG. 1 zeigt ein Blockdiagramm des gesamten Verfahrens.

FIG. 2 zeigt einen Vergleich der gemittelten Ausgangsleistungen Z verschiedener Verfahren mit der Lei-

stung des Original-Geräuschsignals (Beispiel: Mikrofonabstand 12cm, Fahrzeug mit 140km/h). Es wird die zunehmende Geräuschreduktion gezeigt wenn die Verarbeitung mit einem Kanal, mit zwei Kanälen und mit zwei Kanälen mit Nachverarbeitung durchgeführt wird.

Die Mikrofonsignale x und y werden in den Frequenzbereich transformiert (FFT, Fast Fourier-Transformation). Die Segmente sind halb überlappt und werden mit einem Hanning-Fenster gewichtet. Die Segmente sind jeweils N Werte lang und werden um weitere N Nullen erweitert. Die Transformationslänge wird beispielsweise zu  $2N = 512$  gewählt. Es ergeben sich die transformierten Segmente  $X_l(i)$  und  $Y_l(i)$ . Das Ausgangssignal z ergibt sich nach Rücktransformation und der Überlappung der Segmente. l bezeichnet den Blockindex der Segmente, i die diskrete Frequenz ( $i=0,1,2,\dots,2N-1$ ). Die Abtastrate der Signale x und y beträgt z.B. 12kHz.

Im Frequenzbereich wird der Langzeitmittelwert des Betragsspektrums subtrahiert (Spektrale Subtraktion  $H_{SPS}$ ). Das Kurzzeitmittel K und das Langzeitmittel L werden benutzt, um eine erste adaptive Glättungskonstante  $\beta$  zu berechnen. Mit  $\beta$  wird das Störspektrum  $S_{nn}(i)$  geschätzt. Diese adaptive Glättungskonstante ersetzt den sonst üblichen Sprachpausendetektor. l bezeichnet den Blockindex, i die diskrete Frequenz. Als Glättungskonstante  $\beta_0$  wird z.B.  $\beta_0 = 0.03$  verwendet.

$$K_l = \frac{1}{2N} \sum_{i=0}^{2N-1} |X_l(i)|^2 \quad (1)$$

$$\beta_l = g\beta_0 \quad (2)$$

mit

$$g_l = \frac{2L_{l-1}}{L_{l-1} + K_l} \quad (3)$$

$$L_l = (1 - \beta_l)L_{l-1} + \beta_l K_l \quad (4)$$

$$\hat{S}_{nn,l}(i) = (1 - \beta_l)\hat{S}_{nn,l-1}(i) + \beta_l |X_l(i)|^2 \quad (5)$$

Das Störspektrum wird normiert und subtrahiert.

$$|\hat{X}_l(i)| = |X_l(i)| - \frac{\hat{S}_{nn,l}(i)}{|X_l(i)|} \quad (6)$$

$$\hat{X}_l(i) = (1 - \frac{\hat{S}_{nn,l}(i)}{|X_l(i)|^2}) X_l(i) \quad (7)$$

Eine modifizierte Form ergibt sich mit:

$$\hat{X}_l(i) = (1 - \alpha \sqrt{\frac{\hat{S}_{nn,l}(i)}{S_{xx,l}(i)}}) X_l(i); \text{ für } (1 - \alpha \sqrt{\frac{\hat{S}_{nn,l}(i)}{S_{xx,l}(i)}}) X_l(i) < f_0 S_{nn,l}(i) \quad (8)$$

$$\hat{X}_l(i) = f_0 S_{nn,l}(i); \text{ sonst} \quad (9)$$

Für die Leistungsdichte  $S_{xx,l}$  eines Kanales gilt:

$$S_{xx,l}(i) = (1 - \alpha_l)\hat{S}_{xx,l-1}(i) + \alpha_l |X_l(i)|^2 \quad (10)$$

$$\alpha_l = 2 - g_l; \text{ für } 0.5 < 2 - g_l < 2.0 \quad (11)$$

$$\alpha_l = 0.5; \text{ für } 0.5 > 2 - g_l \quad (12)$$

$$\alpha_l = 2; \text{ für } 2 < 2 - g_l \quad (13)$$

$f_0$  wird als "spectral floor" bezeichnet. Es wird ein Teil des Hintergrundgeräuschs zugelassen, um einen natürlich Höreindruck zu erzeugen und um einen Teil der "musical tones" zu maskieren.  $\alpha$  ist ein Überschätzfaktor für das Geräusch und dient der weiteren Reduzierung des Restgeräuschs. Für diese Werte kann z.B.  $f_0 = 0.2$  und  $\alpha = 1.5$  gewählt werden.

Im Gegensatz zu den bekannten Formen der spektralen Subtraktion wird eine zweite adaptive Glättung mit  $\alpha$  dazu benutzt einen weiteren Teil der "musical tones" zu reduzieren, indem die Leistungsdichte  $S_{xx}$  bei Sprache wenig und bei Pause stark geglättet wird.

Für den zweiten Kanal Y gelten die entsprechenden Gleichungen.

Zur Berechnung der linearen Phasenverschiebung zwischen Nutzanteilen in den Kanälen wird das in der nicht vorveröffentlichten Patentanmeldung P 42 43 831 angegebene Verfahren verwendet. Dieses Verfahren fügt sich nahtlos in das erfindungsgemäße Geräuschreduktionsverfahren ein. Die Phasenverschiebung wird

an einer ausgewählten Anzahl der Maximas der Kreuzleistungsdichte geschätzt und die Phasenkorrektur durch Multiplikation im Frequenzbereich mit der Allpaßfunktion  $H_{ALLP}$  erreicht.

$$\hat{X}_l(i) = \hat{X}_l(i) H_{ALLP,i} \quad (14)$$

$$\hat{X}_l(i) = \hat{X}_l(i) (\cos(i * \phi) + j \sin(i * \phi)) \quad (15)$$

Bei mehr als zwei Kanälen wird die Phasenkorrektur für den jeweils weiteren Kanal durchgeführt. Der erste Kanal dient als Referenz.

Durch ein "Beamforming-Verfahren" werden für die Kanäle die Richtungsfilter berechnet. Dabei können als Geräusch verschiedene Fälle betrachtet werden. Es ergeben sich entsprechend der Geräuschsituation verschiedene Richtungsfilter  $H_R$ . Es wird ein Satz dieser Filter ausgewählt, jedoch kann falls im späteren Betrieb der Systemzustand bekannt ist, auf einem bestimmten Satz umgeschaltet werden oder die Filter können ständig adaptiert werden. Als "Beamforming-Verfahren" wird beispielsweise das Gradientenverfahren nach Frost ("An Algorithm for Linearly Constrained Adaptive Array Processing" Proc. IEEE, Vol. 60, No. 8, 1972) oder nach Sondhi und Elko ("Adaptive Optimization of Microphone Arrays under a Nonlinear Constraint" Int. Conf. on ASSP, Tokyo, 1996, S. 981-984) verwendet.

Für die Richtungsfilterung ergibt sich im Frequenzbereich die Multiplikation:

$$\hat{X}_l(i) = \hat{X}_l(i) H_{R,i} \quad (16)$$

Die Addition der Kanäle ergibt mit den Richtungsfiltern die Gesamt-Richtcharakteristik und das Ausgangssignal

$$Z_l(i) = \hat{X}_l(i) + \hat{Y}_l(i) \quad (17)$$

Außerdem führt die Addition der Kanäle zu einer Mittelung und damit Reduzierung der statistischen Fehler der spektralen Subtraktion.

Anschließend wird die Kreuzleistungsdichte der beiden Kanäle mit Hilfe einer Glättungskonstanten (z.B.  $\gamma = 0.3$ ) berechnet.

$$S_{xy,i}(i) = (1 - \gamma) S_{xy,i-1}(i) + \gamma \hat{X}_l(i) \hat{Y}_l^*(i) \quad (18)$$

Die Kreuzleistungsdichte  $S_{xy}$  wird mit der Summe der Leistungsdichten  $S_{xx}$ ,  $S_{yy}$  der einzelnen Kanäle normiert. Es ergibt sich eine modifizierte Kohärenzfunktion:

$$H_{KKF,i}(i) = \frac{S_{xy,i}(i)}{S_{xx,i}(i) + S_{yy,i}(i)}; \text{ für } \frac{S_{xy,i}(i)}{S_{xx,i}(i) + S_{yy,i}(i)} > 0.3 \quad (19)$$

$$H_{KKF,i}(i) = 0.3; \text{ sonst} \quad (20)$$

mit

$$S_{xx,i}(i) = (1 - \gamma) S_{xx,i-1}(i) + \gamma \hat{X}_l(i) \hat{X}_l^*(i) \quad (21)$$

$$S_{yy,i}(i) = (1 - \gamma) S_{yy,i-1}(i) + \gamma \hat{Y}_l(i) \hat{Y}_l^*(i) \quad (22)$$

Für das Ausgangssignal  $Z$  gilt:

$$Z_l(i) = Z_l(i) H_{KKF,i}(i) \quad (23)$$

Werden Richtungsfilter nach dem Verfahren von Sondhi und Elko verwendet, ist ein inverses Filter zur Frequenzgangkorrektur erforderlich. Dieses Filter dient der Anhebung der tieferen Frequenzen, weil der Frequenzgang der Richtungsfilter (für die gewünschte Richtung, Richtung des Sprechers) zu einer Absenkung dieser Frequenzen führt. Dieses Filter  $H_{INV}$  kann auf einfache Weise aus dem berechneten Frequenzgang approximiert werden.

$$Z_l(i) = Z_l(i) H_{INV,i}(i) \quad (24)$$

Wird die Adaption nach dem Verfahren von Frost durchgeführt, ist kein inverses Filter erforderlich, weil der Frequenzgang in Richtung des Sprechers den konstanten Wert 1 hat.

Das erfindungsgemäße Verfahren ist nicht auf Systeme mit zwei Kanälen beschränkt, sondern auf Mehrkanalsysteme (3 und mehr Kanäle) anwendbar.

## Patentansprüche

1. Verfahren zur Geräuschreduktion von zumindest zwei gestörten Sprachkanälen, wobei die gestörten Sprachkanäle zu einem Ausgangskanal zusammengeführt werden, dadurch gekennzeichnet,
  - daß mittels digitaler Richtungsfilter und einer linearen Phasenschätzung für die einzelnen Kanäle eine schwenkbare, akustische Richtkeule erzeugt wird, die der Sprecherbewegung folgt und dadurch die räumlichen Störquellen gedämpft werden,
  - daß in den einzelnen Kanälen in den Sprechpausen die Störung geschätzt wird und durch spektrale Subtraktion die zeitlich stationären Störquellen gedämpft werden,
  - daß anschließend die einzelnen Sprachkanäle addiert werden, und dadurch die statistischen Störungen der spektralen Subtraktion gemittelt werden, und

- daß das Summensignal mit einer modifizierten Kohärenzfunktion nachverarbeitet wird und dadurch die diffusen Stör- und Hallanteile gedämpft werden.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet,

- 5
- daß die spektrale Subtraktion mit zwei adaptiven Glättungskonstanten  $\alpha$ ,  $\beta$  durchgeführt wird,
  - daß mit der ersten adaptiven Glättungskonstante  $\beta$  das Störspektrum  $S_{nn}$  geschätzt wird, und
  - daß mit der zweiten adaptiven Glättungskonstanten  $\alpha$  die Leistungsdichte  $S_{xx}$  der einzelnen Kanäle in den Sprachpausen stark und bei Sprache wenig geglättet wird.

10 3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die linearen Phasenverschiebung von zumindest zwei Signalen über eine bestimmte Anzahl von Maxima der Kreuzleistungsdichte im Frequenzbereich ermittelt wird.

15 4. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Phasenkorrektur, die Richtungsfilterung und eine eventuelle notwendige inverse Filterung im Frequenzbereich durchgeführt werden.

20

25

30

35

40

45

50

55

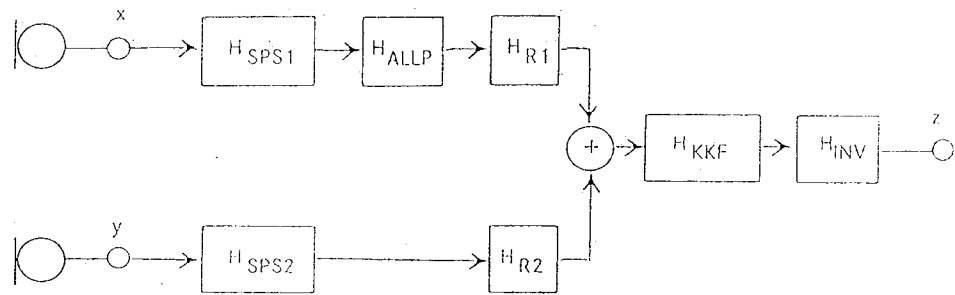


FIG. 1

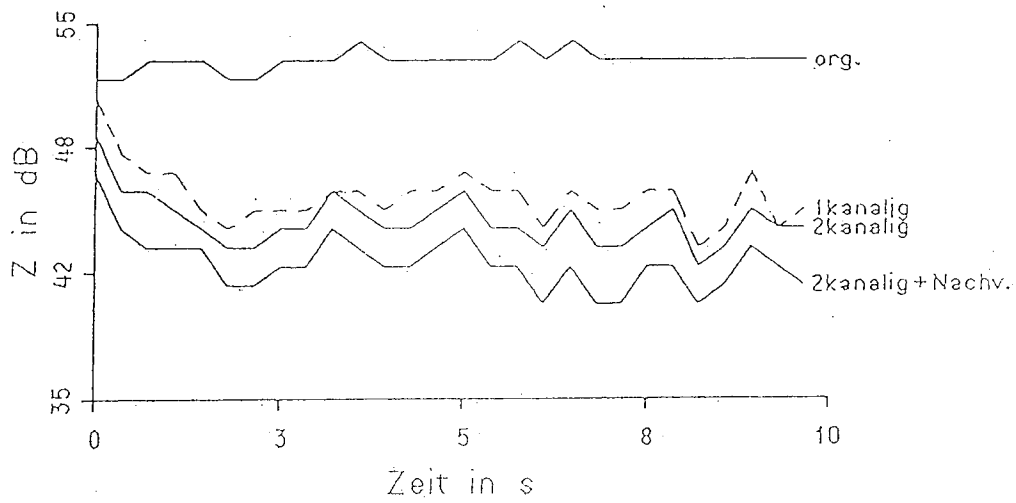


FIG. 2