



(12) **EUROPÄISCHE PATENTANMELDUNG**

(43) Veröffentlichungstag:  
**21.02.2007 Patentblatt 2007/08**

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

(21) Anmeldenummer: **06014433.4**

(22) Anmeldetag: **12.07.2006**

(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 NL PL PT RO SE SI SK TR**  
 Benannte Erstreckungsstaaten:  
**AL BA HR MK YU**

(71) Anmelder: **Micronas GmbH**  
**79108 Freiburg (DE)**

(72) Erfinder: **Fischer, Jörn**  
**79102 Freiburg (DE)**

(30) Priorität: **19.08.2005 DE 102005039621**

(74) Vertreter: **Patentanwälte Westphal, Mussnug & Partner**  
**Am Riettor 5**  
**78048 Villingen-Schwenningen (DE)**

(54) **Verfahren und Vorrichtung zur adaptiven Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System**

(57) Die Erfindung bezieht sich auf eine Vorrichtung bzw. auf ein Verfahren zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System, bei dem  
 - ein Audio-Eingangssignal ( $s(t)$ ) gefiltert wird mittels einer Filterung unter Einsatz eines adaptiven Filters zum Erzeugen eines Vorhersage-Ausgangssignals ( $sv(t)$ ) mit reduziertem Rauschen, wobei das Filtern durchgeführt wird unter Einsatz einer Vielzahl von Koeffizienten ( $c_i(t)$ ;

$c_1 - c_4$ ) zur Bildung einer Vielzahl von Vorhersagefehlern ( $sv_1 - sv_4$ ) und zur Bildung eines Fehlers ( $e$ ) aus der Vielzahl von Vorhersagefehlern ( $sv_1 - sv_4$ ),  
 - wobei mittels einer Vielzahl von Reduktionsparametern ( $k$ ) die Beträge der Koeffizienten ( $c_i(t)$ ;  $c_1 - c_4$ ) kontinuierlich reduziert werden. Nachgeschaltet wird dieser ersten Filterung vorzugsweise eine zweite Filterung mit einer um mehrere Zehner-Potenzen geringeren Lernrate für die Koeffizienten der zweiten Filterung gegenüber einer Lernrate für die Koeffizienten der ersten Filterung.

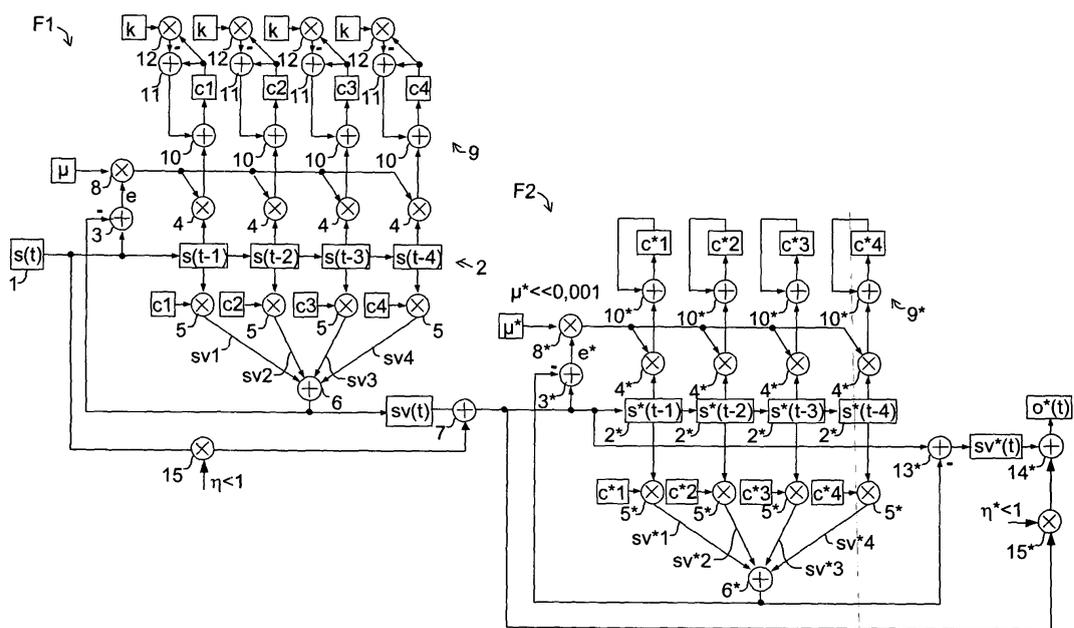


Fig. 1

## Beschreibung

**[0001]** Die Erfindung bezieht sich auf ein Verfahren zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System mit den oberbegrifflichen Merkmalen des Patentanspruchs 1 bzw. auf eine Vorrichtung zum Durchführen eines solchen Verfahrens mit den oberbegrifflichen Merkmalen des Patentanspruchs 18.

**[0002]** In sprachverarbeitenden Systemen, wie z.B. Systemen zur Spracherkennung, Sprachdetektion oder Sprachkomprimierung, mindern Störungen wie Rauschen und nicht zur Sprache gehörende Hintergrundgeräusche die Qualität der Sprachverarbeitung z.B. hinsichtlich der Erkennung oder Kompression der in einem Eingangssignal enthaltenen Sprach- bzw. Sprachsignalanteile. Diese störenden Hintergrundsignale sollen mit möglichst wenig Rechenaufwand eliminiert werden.

**[0003]** Zur Reduktion von Rausch- und Hintergrundsignalen in sprachverarbeitenden Systemen werden Filtervorrichtungen eingesetzt welche mittels mindestens eines Audioeingangs, eines Audioausgangs, eines Speichers und eines Prozessors oder eines feldprogrammierbaren Bausteins oder eines ASIC (Application-Specified-Integrated-Circuit / Anwendungsspezifische integrierte Schaltung) ein Filterverfahren ausführen.

**[0004]** Mittels eines aufwändigen Verfahrens unter Einsatz einer spektralen Subtraktion wird in EP 1080465 und in US 6,820,053 zur Reduktion von Rausch- und Hintergrundsignalen ein Spektrum eines Audiosignals mit Hilfe der Fourier-Transformation berechnet und z. B. ein langsam ansteigender Anteil abgezogen. Durch Rücktransformation in den Zeitbereich wird nachfolgend ein rauschreduziertes Ausgangssignal erhalten. Der Rechenaufwand ist bei diesem Verfahren ist nachteilhaft hoch. Außerdem ist der Speicherplatzverbrauch sehr hoch. Außerdem lassen sich bei der spektralen Subtraktion die verwendeten Parameter zum Teil sehr schlecht an andere Sampleraten anpassen.

**[0005]** Zur Reduktion von Rausch- und Hintergrundsignalen gibt es weitere Verfahren, wie das center clipping (zentrales Klipping), bei dem eine Autokorrelation des Signals gebildet und als Information des Rauschgehalts des Eingangssignals benutzt wird, Verfahren, die gemäß US 5,583,968 oder US 6,820,053 mit neuronalen Netzen arbeiten, welche aufwändig trainiert werden müssen oder Verfahren, gemäß z.B. US 5,500,903 die mit mehreren Mikrofonen arbeiten, um Rauschen und Sprachsignale zu trennen. Mindestens aber wird eine Abschätzung der Rauschamplituden durchgeführt. Der Rechenaufwand einer Fouriertransformation (FFT) ist  $O(n \log(n))$ , der einer Autokorrelation  $O(n^2)$ , der des hier vorgestellten gesamten Verfahrens liegt bei  $O(n)$ .

**[0006]** Allgemein bekannt ist dabei der Einsatz eines FIR-Filters (FIR: Finite Impulse Response / Finite Impulsantwort), das darauf trainiert ist, das Eingangssignal aus z.B. Sprache und Rauschen möglichst gut aus den vergangenen  $n$  Werten vorherzusagen, wobei dies unter Einsatz einer LPC (Linear Predictive Coding / lineare Vorhersage-Kodierung) durchgeführt wird. Die Ausgabewerte des Filters sind diese vorhergesagten Werte. Die Beträge von Koeffizienten  $c(i)$  eines solchen Filters steigen bei Rauschsignalen im Mittel langsamer als bei Sprachsignalen, wobei die Koeffizienten berechnet werden gemäß

$$c_i(t+1) = c_i(t) + \mu \cdot e \cdot s(t-i) \quad (1)$$

mit  $\mu \ll 1$ , z.B.  $\mu = 0,01$  als einer Lernrate,  $s(t)$  als einem Audio-Eingangssignal zur Zeit  $t$ ,  $e = s(t) - sv(t)$  als einem Fehler aus einer Differenz aller einzelner Vorhersagefehler vom Audio-Eingangssignal,  $sv(t)$  als Ausgangssignal aus einer Summe der Terme  $c_i(t-1) \cdot s(t-i)$ , d.h. der einzelnen Vorhersagefehler über alle  $i$  von 1 bis  $N$ ,  $N$  als Anzahl der Koeffizienten und  $c_i(t)$  als einem individuellen Koeffizienten mit einem Parameter  $i$  zur Zeit  $t$ .

**[0007]** Die Aufgabe der Erfindung besteht darin, ein Verfahren zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System bzw. eine Vorrichtung zum Durchführen eines solchen Verfahrens hinsichtlich der Anwendbarkeit zu verbessern, insbesondere flexibler einsetzbar zu machen.

**[0008]** Diese Aufgabe wird durch ein Verfahren zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System mit den Merkmalen des Patentanspruchs 1 bzw. auf eine Vorrichtung zum Durchführen eines solchen Verfahrens mit den Merkmalen des Patentanspruchs 18 gelöst. Vorteilhafte Ausgestaltungen sind Gegenstand von abhängigen Ansprüchen.

**[0009]** Bevorzugt wird demgemäß ein Verfahren zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System, bei dem ein Audio-Eingangssignal gefiltert wird mittels einer Filterung unter Einsatz eines adaptiven Filters zum Erzeugen eines Vorhersage-Ausgangssignals mit reduziertem Rauschen, wobei das Filtern durchgeführt wird unter Einsatz einer Vielzahl von Koeffizienten zur Bildung einer Vielzahl von Vorhersagefehlern und zur Bildung eines Fehlers aus der Vielzahl von Vorhersagefehlern, wobei mittels einer Vielzahl von Reduktionsparametern die Beträge der Koeffizienten kontinuierlich reduziert werden.

**[0010]** Bevorzugt wird insbesondere ein Verfahren, bei dem die kontinuierliche Reduktion der Koeffizienten dadurch erzeugt wird, dass die Koeffizienten mit einem Faktor kleiner 1 multipliziert werden, insbesondere mit einem Faktor

## EP 1 755 110 A2

zwischen 0,8 und 1,0 multipliziert werden.

**[0011]** Bevorzugt wird insbesondere ein Verfahren, bei dem die Koeffizienten  $c_i(t)$  berechnet werden gemäß

$$c_i(t+1) = c_i(t) + \mu \cdot e \cdot s(t-i) - kc_i(t)$$

mit

- $k$  mit  $0 > k \ll 1$ , insbesondere  $k \leq 0,0001$ , als einem Reduktionsparameter,
- $\mu \ll 1$ , insbesondere  $\mu \leq 0,01$ , als einer Lernrate,
- $s(t)$  als einem Audio-Eingangssignal zu einer Zeit  $t$ ,
- $e$  als einem Fehler aus einer Differenz aller einzelner Vorhersagefehler ( $sv_1 - sv_4$ ) vom Audio-Eingangssignal  $s(t)$ ,
- $sv(t)$  als dem Vorhersage-Ausgangssignal aus einer Summe aller einzelnen Vorhersagefehler, mit  $N$  als Anzahl der Koeffizienten  $c_i(t)$  und
- $c_i(t)$  als individuellem Koeffizient mit einem Index  $i$  zur Zeit  $t$ .

**[0012]** Bevorzugt wird insbesondere ein solches Verfahren, bei dem die Koeffizienten berechnet werden gemäß

$$c_i(t+1) = c_i(t) + \mu \cdot e \cdot s(t-i) - kc_i(t)$$

mit

$$e = S(t) - sv(t)$$

und

$$sv(t) = \sum_{i=1 \dots N} c_i(t-1) \cdot s(t-i).$$

**[0013]** Bevorzugt wird insbesondere ein Verfahren, bei dem das Vorhersage-Ausgangssignal als eine Vorhersage des Audio-Eingangssignals mit reduziertem Rauschen als Eingangssignal für eine nachfolgende zweite Filterung verwendet wird zum Erzeugen einer zweiten Vorhersage. Bevorzugt wird insbesondere ein solches Verfahren, bei dem die zweite Filterung mittels einer Vorhersagefilterung mit einer zweiten, insbesondere für sich bekannten Filterung mit einem Satz von zweiten Koeffizienten durchgeführt wird, wobei eine Lernrate zum Anpassen der Koeffizienten um einige Zehnerpotenzen kleiner gewählt wird als eine Lernrate der ersten Filterung.

**[0014]** Bevorzugt wird insbesondere ein Verfahren, bei dem die zweite Vorhersage dann vom Vorhersage-Ausgangssignal abgezogen wird, um lang anhaltende Hintergrundgeräusche zu eliminieren.

**[0015]** Bevorzugt wird insbesondere ein Verfahren, bei dem eine Lernregel zum Bestimmen der weiteren Koeffizienten unsymmetrisch gestaltet wird, so dass der Betrag der weiteren Koeffizienten im Betrag stärker fallen als steigen und schnell auf Null absinken kann aber nur mit kleiner Steigung ansteigt.

**[0016]** Bevorzugt wird insbesondere ein Verfahren, bei dem statt des Audio-Eingangssignals zum Bestimmen individueller Vorhersagefehler nur dessen Vorzeichen benutzt wird, um kleine Signale nicht zu benachteiligen.

**[0017]** Bevorzugt wird insbesondere ein Verfahren, bei dem die Koeffizienten begrenzt werden zum Vermeiden eines Abdriftens der Koeffizienten, insbesondere auf einen Bereich von z.B. -4 ... 4, wenn das Audio-Eingangssignal von -1 ... 1 normiert ist.

**[0018]** Bevorzugt wird insbesondere ein Verfahren, bei dem ein Maximum eines Sprachsignalanteils des Audio-Eingangssignals detektiert wird und das Ausgangssignal wieder auf dieses Maximum insbesondere schleppend normiert wird.

**[0019]** Bevorzugt wird insbesondere ein Verfahren, bei dem das Ausgabesignal der ersten und/oder der zweiten Filterung im Verhältnis zu deren Eingabesignal insbesondere gleichzeitig als ein Maß für das Vorhandensein von Sprache im Eingabesignal verwendet wird.

**[0020]** Bevorzugt wird insbesondere ein Verfahren, bei dem für die erste und/oder die zweite Filterung ein Filter verwendet wird, das mittels einer LMS-Adaption (Least Mean Squares Adaption) eine Fehlervorhersage durchführt.

Bevorzugt wird insbesondere ein Verfahren, bei dem für die erste und/oder die zweite Filterung ein FIR-Filter verwendet wird.

**[0021]** Bevorzugt wird insbesondere ein Verfahren, bei dem mit dem Vorhersage-Ausgangssignal eine sigmoide Funktion multipliziert wird zum Vermeiden eines Übersteuerns des Signals im Fall einer schlechten Vorhersage.

**[0022]** Bevorzugt wird insbesondere ein Verfahren, bei welchem dem Vorhersage-Ausgangssignal als Originalsignals das Audio-Eingangssignal zugemischt wird zum Erzeugen eines natürlicheren Klangs.

**[0023]** Bevorzugt wird insbesondere ein Verfahren, bei dem zum Durchführen des Verfahrens ein feldprogrammierbarer Baustein oder eine ASIC (Application-Specified-Integrated-Circuit) entsprechend programmiert wird.

**[0024]** Bevorzugt wird demgemäss eine Vorrichtung, insbesondere Vorrichtung zum Durchführen eines Verfahrens, zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System, mit einem Audio-Eingang zum Eingeben eines Audio-Eingangssignals einem adaptiven Filter zum Filtern Audio-Eingangssignals zum Erzeugen eines Vorhersage-Ausgangssignals mit reduziertem Rauschen, mit einem Speicher zum Speichern einer Vielzahl von Koeffizienten für das Filter, wobei das Filter ausgebildet oder geschaltet ist zur Bildung einer Vielzahl von Vorhersagefehlern und zur Bildung eines Fehlers aus der Vielzahl von Vorhersagefehlern, wobei eine Koeffizienten-Bereitstellungsanordnung ausgebildet oder geschaltet ist mittels zumindest einem Reduktionsparameter die Beträge der Koeffizienten kontinuierlich zu reduzieren.

**[0025]** Bevorzugt wird insbesondere eine Vorrichtung, bei der die Koeffizienten-Bereitstellungsanordnung zum Multiplizieren der Koeffizienten mit dem Reduktionsparameter als einem Faktor kleiner 1, insbesondere mit einem Faktor zwischen 0,8 und 1,0 ausgebildet oder geschaltet ist.

**[0026]** Bevorzugt wird insbesondere eine Vorrichtung, bei welcher einer ersten Filterstufe mit dem Filter als erstem Filter eine zweite Filterstufe mit einem zweiten Filter nachgeschaltet ist zum Zuführen des Vorhersage-Ausgangssignals als eine Vorhersage des Audio-Eingangssignals mit reduziertem Rauschen als Eingangssignal für den zweiten Filter zum Erzeugen einer zweiten Vorhersage.

**[0027]** Bevorzugt wird insbesondere eine Vorrichtung mit einer Subtraktionsschaltung zum Abziehen einer Summe aus Fehler-Vorhersagen der zweiten Filterung vom Vorhersage-Ausgangssignal zum Erzeugen der Vorhersage.

**[0028]** Bevorzugt wird insbesondere eine Vorrichtung, bei der das zweite Filter durch ein LMS-Adaptions-Filter zum Durchführen einer Fehlervorhersage ausgebildet oder geschaltet ist.

**[0029]** Bevorzugt wird insbesondere eine Vorrichtung, bei der das erste Filter und/oder das zweite Filter durch ein FIR-Filter zum Durchführen einer Signalvorhersage ausgebildet oder geschaltet ist.

**[0030]** Bevorzugt wird insbesondere eine Vorrichtung, welche durch einen feldprogrammierbaren Baustein oder einen ASIC ausgebildet ist.

**[0031]** Bevorzugt wird insbesondere eine Vorrichtung mit einem Multiplizierer zum Wichten des optional zeitlich verzögerten Audio-Eingangssignal oder zum Wichten des Vorhersage-Ausgangssignals mit einem Wichtungsfaktor kleiner Eins, insbesondere etwa 0,1 und einem Addierer zum Addieren des gewichteten Signals auf das Vorhersage-Ausgangssignal oder auf die Vorhersage zum Erzeugen eines rauschreduzierten Audio-Ausgangssignals.

**[0032]** Gegenüber EP 1080465 und US 6,820,053 ist der Rechenaufwand bei dem hier bevorzugten Verfahren um ein Vielfaches geringer. Außerdem ist der Speicherplatzverbrauch um ein Vielfaches geringer. Außerdem entfällt das Problem der sehr schlechten Anpassung der verwendeten Parameter an andere Sampleraten wie bei der spektralen Subtraktion.

**[0033]** Im Vergleich zu den verschiedenen bekannten Verfahren ist der Rechenaufwand sehr viel geringer. Während bei einer Fouriertransformation der Rechenaufwand bei  $O(n \log(n))$  liegt und der Rechenaufwand einer Autokorrelation bei  $O(n^2)$  liegt, liegt der Rechenaufwand des hier bevorzugten Verfahrens gesamten Verfahrens beider Filterstufen zusammen nur bei  $O(n)$ , wobei  $n$  eine Anzahl abgetasteter Samples (Stützstellen) des Eingangssignals und  $O$  eine allgemeine Funktion des Filteraufwands ist.

**[0034]** Durch die besonders bevorzugte Filteranordnung ergibt sich eine Vielzahl an Vorteilen. Ein Sprachsignal wird nur um ein einziges Sample verzögert. Eine Adaption ist für rauschen instantan und für lang anhaltende Hintergrundgeräusche ist die Adaption vorzugsweise etwa 0,2 s bis 5,0 s verzögert.

**[0035]** Das Verfahren ist wesentlich weniger rechenaufwändig als übliche Verfahren. Schon mit insbesondere nur vier Koeffizienten erhält man respektable Ergebnisse, so dass nur vier Multiplikationen und vier Additionen für die Vorhersage eines Samples berechnet werden müssen und nur vier bis fünf weitere Operationen für die Adaption der Filterkoeffizienten erforderlich sind.

**[0036]** Hinzu kommt ein geringerer Speicherplatzverbrauch als für übliche Verfahren, wie z.B. der spektralen Subtraktion.

Ermöglicht wird ein einfaches Justieren der Parameter auch bei verschiedenen Sampleraten. Außerdem kann die Stärke der Filterung für Rauschen und für lang anhaltende Hintergrundsignale separat eingestellt werden.

**[0037]** Ein Ausführungsbeispiel wird nachfolgend anhand der Zeichnung näher erläutert. Es zeigen:

Fig. 1 eine bevorzugte Filteranordnung zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverar-

beitenden System mit zwei hintereinander geschalteten Filterstufen,

Fig. 2 vergrößert dargestellt die erste der beiden Filterstufen und

Fig. 3 vergrößert dargestellt die zweite der beiden Filterstufen.

5 **[0038]** Wie aus Fig. 1 ersichtlich, besteht das besonders bevorzugte Verfahren aus zwei adaptiven Filtern F1, F2, welche als eine erste und eine zweite Filterstufe hintereinander geschaltet sind. Eigenständig vorteilhaft ist jedoch bereits der Einsatz nur der ersten Filterstufe.

**[0039]** In die besonders bevorzugte Schaltungsanordnung wird über einen Audio-Eingang 1 ein Audio-Eingangssignal  $s(t)$  eingegeben. Das Audio-Eingangssignal wird einer Gruppe von Verzögerungsgliedern 2 angelegt, welche z. B. als Puffer ausgebildet sind und den jeweils anliegenden Wert des Audio-Eingangssignals  $s(t)$  um jeweils einen Takt verzögern. Außerdem wird das Audio-Eingangssignal  $s(t)$  einem ersten Addierer 3 zugeführt. Die mittels der Verzögerungsglieder 2 verzögerten Werte  $s(t-1) - s(t-4)$  werden von dem jeweiligen Verzögerungsglied 2 aus jeweils dem nächsten der Verzögerungsglieder 2 sowie jeweils zwei entsprechenden Multiplizierern zweier Gruppen aus Multiplizierern 4 angelegt. Der Gruppe zweiter Multiplizierer 5 wird an einen weiteren Multiplikationseingang jeweils ein Koeffizient  $c1 - c4$  als Filterkoeffizient eines adaptiven Filters angelegt. Die Multiplikationsergebnisse der Gruppe zweiter Multiplizierer 5 werden als individuelle Vorhersagefehler  $sv1 - sv4$  an einen zweiten Addierer 6 ausgegeben. Eine zeitliche Abfolge der Additionswerte des zweiten Addierers 6 bildet ein Vorhersage-Ausgangssignal  $sv(t)$ .

**[0040]** Die Abfolge der Werte des Vorhersage-Ausgangssignals  $sv(t)$  werden gemäß einer ersten vorteilhaften Ausführungsform direkt ausgegeben, um ein Ausgangssignal  $o(t)$  zu bilden (Fig. 2).

20 **[0041]** Die Abfolge der Werte des Vorhersage-Ausgangssignals  $sv(t)$  werden außerdem dem als Subtraktionsschaltung ausgebildeten ersten Addierer 3 an einem Subtraktionseingang angelegt, um diese Werte vom momentanen späteren Wert des Audio-Eingangssignals  $s(t)$  abzuziehen. Das Subtraktionsergebnis des ersten Addierers 3 bildet einen Fehler  $e$  aus einer entsprechenden Abfolge einzelner Fehlerwerte. Dieser Fehler  $e$  wird einem dritten Multiplizierer 8 angelegt, an dessen zweitem Multiplikationseingang ein Wert einer Lernrate  $\mu$  mit vorzugsweise  $\mu \approx 0,01$  angelegt. Das Multiplikationsergebnis wird den Eingängen der Gruppe erster Multiplizierer 4 zur Multiplikation mit den verzögerten Werten  $s(t-1) - s(t-4)$  angelegt.

25 **[0042]** Die Multiplikationsergebnisse der Gruppe erster Multiplizierer 4 werden einer Gruppe aus dritten Addierern 10 zugeführt, welche einen Eingang einer Koeffizienten-Bereitstellungsanordnung 9 ausbilden. Die Ausgangswerte der Gruppe dritter Addierer 10 bilden die Koeffizienten  $c1 - c3$  aus, welche den entsprechenden Multiplizierern 5 der Gruppe zweiter Multiplizierer 5 angelegt werden. Außerdem werden diese Koeffizienten  $c1 - c4$  jeweils einem Addierer 11 einer Gruppe vierter Addierer 11 und jeweils einem Multiplizierer 12 einer Gruppe vierter Multiplizierer 12 angelegt. An die Gruppe vierter Multiplizierer 12 wird an einen Multiplikationseingang ein Reduktionsparameter  $k$  angelegt, wobei der Wert des Reduktionsparameters  $k$  beispielsweise 0,0001 beträgt. Durch den Reduktionsparameter  $k$  wird entsprechend der jeweilige Wert der Koeffizienten  $c1 - c4$  um diesen Faktor reduziert. Das entsprechende Multiplikationsergebnis der vierten Multiplizierer 12 wird dem jeweiligen der als Subtraktionsschaltung ausgebildeten vierten Addierer 11, dem zuvor der entsprechende Koeffizient  $c1 - c4$  angelegt wurde, an einem Subtraktionseingang angelegt. Der Ausgangswert der jeweiligen Addierer 11 der vierten Gruppe Addierer 11 wird einem weiteren Eingang des entsprechenden dritten Addierers der Gruppe dritter Addierer 10 angelegt. Dabei wird von der Gruppe dritter Addierer 10 der jeweilige Additionswert der Gruppe vierter Addierer 11 auf den jeweils angelegten und verzögerten Audio-Signaleingangswert  $s(t-1) - s(t-4)$  aufaddiert, um die Koeffizienten zu lernen.

30 **[0043]** Mittels eines Addierers 7 kann dem Vorhersage-Ausgangssignal  $sv(t)$  zur Bildung des Ausgangssignals  $o(t)$  optional ein gewichteter Wert aufaddiert werden, der direkt vom momentanen oder optional von einem entsprechend verzögerten Wert des Audio-Eingangssignals  $s(t)$  gebildet wird. Der gewichtete Wert wird durch einen Wichtungs-Multiplizierer 15 bereitgestellt, der das Eingangssignal  $s(t)$  mit einem Faktor  $\eta < 1$ , insbesondere  $\eta \approx 0,1$  multipliziert.

35 **[0044]** Vorzugsweise wird das Vorhersage-Ausgangssignal  $sv(t)$  bzw. das Ausgangssignal  $o(t)$  nicht als endgültiges Ausgangssignal ausgegeben sondern als Eingangssignal für eine zweite Filterstufe mit dem zweiten Filter F2 bereitgestellt.

40 **[0045]** Wie dies in Fig. 3 dargestellt ist, handelt es sich bei dem zweiten Filter F2 wiederum um eine adaptive Filteranordnung, wobei deren Aufbau vorzugsweise im Wesentlichen gleich dem Aufbau der ersten Filterstufe ist. Nachfolgend werden daher lediglich Unterschiede zu der ersten Filterstufe beschrieben. Die jeweiligen Komponenten und Signale bzw. Werte sind mit einem Stern zur Unterscheidung entsprechender Komponenten und Signale bzw. Werte der ersten Filterstufe gekennzeichnet.

45 **[0046]** Unterschiedlich ist die Erzeugung der Koeffizienten  $c^*1 - c^*4$  in einer gegenüber der ersten Filterstufe modifizierten Koeffizienten-Bereitstellungseinrichtung 9\*. Die Koeffizienten  $c^*1 - c^*4$  werden in für sich bekannter Art und Weise eines z. B. adaptiven FIR-Filter ohne eine Multiplikation mit einem Reduktionsparameter  $k$  gebildet. Ein weiterer Unterschied gegenüber sowohl der ersten Filterstufe des ersten Filters F1 als auch einem herkömmlichen FIR-Filter besteht darin, dass der Wert einer Lernrate  $\mu^*$  für den zweiten Filter F2 kleiner, insbesondere deutlich kleiner als der Wert der Lernrate  $\mu$  des ersten Filters F1 gewählt wird.

**[0047]** Das Ausgangsergebnis des zweiten Filters F2 wird entsprechend durch einen zweiten Addierer 6\* des zweiten Filters F2 bereitgestellt und dem Eingangssignal bzw. dem entsprechenden Eingangswert des Eingangssignals sv(t) des zweiten Filters F2 mittels eines fünften Addierers 13\* aufaddiert bzw. vorzugsweise davon subtrahiert im Fall eines bevorzugt als Subtraktionsschaltung ausgebildeten Addierers 6\*. Das Ausgangsergebnis des fünften Addierers 13\* bildet eine zweite Vorhersage sv\*(t) als ein zweites Vorhersage-Ausgangssignal aus. Vorzugsweise werden die Werte der Vorhersage sv\*(t) mittels eines sechsten Addierers 14\* dem optional zeitlich verzögerten und gewichteten Audio-Eingangssignal s(t) bzw. sv(t) aufaddiert zum Erzeugen eines rauschreduzierten Audio-Ausgangssignals o\*(t). Zur Wichtung dient eine Multiplikation des Audio-Eingangssignals s(t) mit einem Wichtungsfaktor  $\eta^* < 1$ , insbesondere  $\eta \approx 0,1$  in einem Multiplizierer 15\*, der dem sechsten Addierer 14\* vorgeschaltet ist. Zum Steuern der Verfahrensschritte weist die Anordnung in üblicher Art und Weise weitere Komponenten auf oder ist an weitere Komponenten wie z.B. einen Prozessor für Steuerfunktionen und einen Taktgeber zum Bereitstellen eines Taktsignals angeschlossen. Zum Speichern der Koeffizienten c1 - c4, c\*1 - c\*4 und ggfs. weiterer Werte weist die Anordnung einen Speicher auf oder kann auf einen Speicher zugreifen.

**[0048]** Das erste Filter F1 reduziert das Rauschen über den gesamten wahrgenommenen Frequenzbereich. Dabei wird ein modifiziertes adaptives FIR-Filter, darauf trainiert, das Audio-Eingangssignal s(t), das z.B. Sprache und Rauschen enthält, möglichst gut aus den vergangenen n Werten vorherzusagen. Die Ausgabe sind die vorhergesagten Werte als das Vorhersage-Ausgangssignal sv(t). Die Beträge der allgemeinen Koeffizienten ci(t) mit gemäß Fig. 1 einem Index i = 1, 2, 3, 4 und entsprechend den Koeffizienten C1 - C4 eines solchen ersten Filters F1 steigen bei Rauschsignalen langsamer als bei Sprachsignalen.

**[0049]** Das Filtern erfolgt in Analogie zur LPC. Statt der einer Delta-Regel oder einem LMS- Lernschritt gemäß dem Stand der Technik wird nun ein modifiziertes Filterverfahren eingesetzt, bei dem die Koeffizienten ci(t) allgemein gemäß einer neuen Lernregel berechnet werden gemäß

$$c_i(t+1) = c_i(t) + \mu \cdot e \cdot s(t-i) - kc_i(t) \quad (2)$$

mit

$$e = S(t) - sv(t), \quad (3)$$

$$sv(t) = \sum_{i=1 \dots N} c_i(t-1) \cdot s(t-i) \quad \text{und} \quad (4)$$

und mit k mit  $0 > k \ll 1$ , z.B.  $k = 0,0001$ , als einem Reduktionsparameter, mit  $\mu \ll 1$ , z.B.  $\mu = 0,01$ , als einer Lernrate, mit s(t) als einem Audio-Eingangssignal zur Zeit t, mit e als einem Fehler aus einer Differenz aller einzelner Vorhersagefehler vom Audio-Eingangssignal, mit sv(t) als einem Vorhersage-Ausgangssignal aus einer Summe der Koeffizienten multipliziert mit den zugehörigen verzögerten Signalen, mit N als Anzahl der Koeffizienten ci(t) und mit ci(t) als individuellem Koeffizient mit einem Parameter bzw. Index i zur Zeit t.

**[0050]** Gemäß der Lernregel unter Einsatz des Reduktionsparameters k werden die Beträge der Koeffizienten ci(t) kontinuierlich reduziert, was bei Rauschsignalen zu kleineren vorhergesagten Amplituden führt als bei Sprachsignalen. Dabei wird mit dem Reduktionsparameter k festgelegt, wie stark das Rauschen unterdrückt werden soll.

**[0051]** Das zweite Filter F2 reduziert lang anhaltende Hintergrundgeräusche. Dabei wird ausgenutzt, dass die Energie von Sprachsignalanteilen im Audio-Eingangssignal s(t) in einzelnen Frequenzbändern immer wieder auf Null abfällt, wohingegen lang anhaltende Töne eher eine gleichbleibende Energie im Frequenzband haben. Ein adaptives FIR-Filter mit extrem kleiner Lernrate von z.B.  $\mu = 0,000001$  wird nun für eine Vorhersage mittels insbesondere LPC so langsam adaptiert, dass der Sprachsignalanteil im Audio-Eingangssignal s(t) mit sehr viel geringerer Amplitude vorhergesagt wird als lang anhaltende Signale. Abschließend wird die derart in dem zweiten Filter F2 erlangte Vorhersage sv\*(t) vom Eingangssignal s(t) abgezogen, so dass die lang anhaltenden Signale aus dem Eingangssignal s(t) eliminiert oder zumindest stark reduziert werden.

**[0052]** Das erste und das zweite Filter F1, F2 wirken besonders effizient, wenn sie hintereinander auf das Eingangssignal s(t) ausgeführt werden, wie dies in Fig. 1 dargestellt ist. Dabei wird erst das erste Filter F1 ausgeführt und dessen Ausgangs- bzw. Vorhersage-Ausgangssignal sv(t) als Eingangssignal an das zweite Filter F2 zur weiteren zusätzlichen Filterung geleitet.

**[0053]** Fig. 1 zeigt schematisch einen Amplitudenverlauf a über der Zeit t eines beispielhaften Eingangssignal s(t) im

Zeitbereich vor und nach der Filterung durch das erste Filter F1 zur Rauschunterdrückung. Während das Eingangssignal  $s(t)$  Sprache und Rauschen enthält, enthält das Vorhersage-Ausgangssignal  $sv(t)$  des ersten Filters F1 Sprache und ein demgegenüber reduziertes Rauschen.

[0054] Fig. 2 zeigt schematisch einen Amplitudenverlauf  $a$  über der Zeit  $t$  eines beispielhaften Eingangssignals  $s(t)$  bzw. des Vorhersage-Ausgangssignals  $sv(t)$  im Frequenzbereich vor und nach der Filterung durch das zweite Filter F2 zur Unterdrückung von lang anhaltenden Hintergrundgeräuschen. Dabei entspricht die x-Achse der Zeit  $t$ , die y-Achse einer Frequenz  $f$  und eine Helligkeit entspricht einer Amplitude. Erkennbar ist ein Spektrum eines markanten 2kHz-Ton im Hintergrund vor dem zweiten Filter F2 gegenüber einem Spektrum mit reduziertem 2kHz-Ton nach dem zweiten Filter F2.

[0055] Anstelle einer kontinuierlichen Reduktion der Koeffizienten  $C1 - C4$  gemäß Formel (2) kann die Reduktion der Koeffizienten  $c_i(t)$  alternativ oder zusätzlich auch dadurch erzeugt werden, dass die Koeffizienten  $c_i(t)$  mit einem festen oder variablen Faktor zwischen insbesondere 0,8 und 1,0 multipliziert werden.

[0056] Vorteilhaft ist ein Verfahren bzw. eine Vorrichtung bei der nach dem Einsatz des ersten Filters F1 mit dessen Vorhersage-Ausgangssignal  $sv(t)$  eine sigmoide Funktion, z.B. ein Tangenshyperbolikus, multipliziert wird, die bei einer schlechten Vorhersage ein Übersteuern des Signals vermeidet.

[0057] Vorteilhaft ist ein Verfahren bzw. eine Vorrichtung, wenn dem Vorhersage-Ausgangssignal ( $sv(t)$ ) als Originalsignals das Audio-Eingangssignal ( $s(t)$ ) zugemischt wird zum Erzeugen eines natürlicheren Klangs.

[0058] Anstelle eines einzigen Reduktionsparameters  $k$  für alle Koeffizienten  $c1 - c4$  können auch mehrere Reduktionsparameter für die verschiedenen Koeffizienten  $c1 - c4$  individuell festgelegt oder bestimmt werden. Insbesondere können der oder die Reduktionsparameter  $k$  auch abhängig von z.B. dem empfangenen Audio-Eingangssignal variiert werden.

## Patentansprüche

1. Verfahren zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System, bei dem

- ein Audio-Eingangssignal ( $s(t)$ ) gefiltert wird mittels einer Filterung unter Einsatz eines adaptiven Filters zum Erzeugen eines Vorhersage-Ausgangssignals ( $sv(t)$ ) mit reduziertem Rauschen,
- wobei das Filtern durchgeführt wird unter Einsatz einer Vielzahl von Koeffizienten ( $c_i(t)$ ;  $c1 - c4$ ) zur Bildung einer Vielzahl von Vorhersagefehlern ( $sv1 - sv4$ ) und zur Bildung eines Fehlers ( $e$ ) aus der Vielzahl von Vorhersagefehlern ( $sv1 - sv4$ ),

### dadurch gekennzeichnet, dass

- mittels einer Vielzahl von Reduktionsparametern ( $k$ ) die Beträge der Koeffizienten ( $c_i(t)$ ;  $c1 - c4$ ) fortlaufend reduziert werden.

2. Verfahren nach Anspruch 1, bei dem die kontinuierliche Reduktion der Koeffizienten ( $c_i(t)$ ) dadurch erzeugt wird, dass die Koeffizienten ( $c_i(t)$ ) mit einem Faktor kleiner 1 multipliziert werden, insbesondere mit einem Faktor zwischen 0,8 und 1,0 multipliziert werden.

3. Verfahren nach Anspruch 1 oder 2, bei dem die Koeffizienten ( $c_i(t)$ ) berechnet werden gemäß

$$c_i(t+1) = c_i(t) + \mu \cdot e \cdot s(t-i) - kc_i(t)$$

mit

- $k$  mit  $0 > k \ll 1$ , insbesondere  $k \leq 0,0001$ , als einem Reduktionsparameter,
- $\mu \ll 1$ , insbesondere  $\mu \leq 0,01$ , als einer Lernrate,
- $s(t)$  als einem Audio-Eingangssignal zu einer Zeit  $t$ ,
- $e$  als einem Fehler aus einer Differenz aller einzelner Vorhersagefehler ( $sv1 - sv4$ ) vom Audio-Eingangssignal  $s(t)$ ,
- $sv(t)$  als dem Vorhersage-Ausgangssignal aus einer Summe aller einzelnen Vorhersagefehler, mit  $N$  als Anzahl der Koeffizienten  $c_i(t)$  und

## EP 1 755 110 A2

-  $c_i(t)$  als individuellem Koeffizient mit einem Index  $i$  zur Zeit  $t$ .

4. Verfahren nach Anspruch 3, bei dem die Koeffizienten ( $c_i(t)$ ) berechnet werden gemäß

$$\cancel{c_i(t+1) = c_i(t) + \mu \cdot e \cdot s(t-i) - kc_i(t)}$$

mit

$$e = S(t) - sv(t)$$

und

$$sv(t) = \sum_{i=1 \dots N} c_i(t-1) \cdot s(t-i).$$

5. Verfahren nach einem vorstehenden Anspruch, bei dem das Vorhersage-Ausgangssignal ( $sv(t)$ ) als eine Vorhersage des Audio-Eingangssignals mit reduziertem Rauschen als Eingangssignal für eine nachfolgende zweite Filterung (F2) verwendet wird zum Erzeugen einer zweiten Vorhersage ( $sv^*(t)$ ).
6. Verfahren nach Anspruch 5, bei dem die zweite Filterung (F2) mittels einer Vorhersagefilterung mit einer zweiten, insbesondere für sich bekannten Filterung mit einem Satz von zweiten Koeffizienten ( $c_i(t)$ ,  $c^*1 - c^*4$ ) durchgeführt wird, wobei eine Lernrate ( $\mu^*$ ) zum Anpassen der Koeffizienten um einige Zehnerpotenzen kleiner gewählt wird als eine Lernrate ( $\mu$ ) der ersten Filterung (F1).
7. Verfahren nach Anspruch 5 oder 6, bei dem die zweite Vorhersage ( $sv^*(t)$ ) dann vom Vorhersage-Ausgangssignal ( $sv(t)$ ) abgezogen wird.
8. Verfahren nach einem der Ansprüche 5 bis 7, bei dem eine Lernregel zum Bestimmen der weiteren Koeffizienten ( $c_i^*(t)$ ;  $c^*1 - c^*4$ ) unsymmetrisch gestaltet wird, so dass der Betrag der weiteren Koeffizienten ( $c_i^*(t)$ ;  $c^*1 - c^*4$ ) im Betrag stärker fallen als steigen und schnell auf Null absinken kann aber nur mit kleiner Steigung ansteigt.
9. Verfahren nach einem vorstehenden Anspruch, bei dem statt des Audio-Eingangssignals ( $S(t)$ ) zum Bestimmen individueller Vorhersagefehler ( $sv1 - sv4$ ) nur dessen Vorzeichen benutzt wird.
10. Verfahren nach einem vorstehenden Anspruch, bei dem die Koeffizienten ( $c_i(t)$ ;  $c1 - c4$ ) begrenzt werden zum Vermeiden eines Abdriftens der Koeffizienten, insbesondere von  $-4 \dots 4$ , wenn das Audio-Eingangssignal von  $-1 \dots 1$  normiert ist.
11. Verfahren nach einem vorstehenden Anspruch, bei dem ein Maximum eines Sprachsignalanteils des Audio-Eingangssignals ( $s(t)$ ) detektiert wird und das Ausgangssignal ( $o(t)$ ) wieder auf dieses Maximum normiert wird.
12. Verfahren nach einem vorstehenden Anspruch, bei dem das Ausgabesignal ( $sv(t)$ ;  $sv^*(t)$ ) der ersten und/oder der zweiten Filterung im Verhältnis zu deren Eingabesignal ( $s(t)$ ;  $sv(t)$ ) als ein Maß für das Vorhandensein von Sprache im Eingabesignal verwendet wird.
13. Verfahren nach einem vorstehenden Anspruch, bei dem für die erste und/oder die zweite Filterung ein Filter verwendet wird, das mittels einer LMS-Adaption (Least Mean Squares Adaption) eine Fehlervorhersage durchführt.
14. Verfahren nach einem vorstehenden Anspruch, bei dem für die erste und/oder die zweite Filterung ein FIR-Filter verwendet wird.
15. Verfahren nach einem vorstehenden Anspruch, bei dem mit dem Vorhersage-Ausgangssignal ( $sv(t)$ ) eine sigmoide Funktion multipliziert wird zum Vermeiden eines Übersteuerns des Signals im Fall einer schlechten Vorhersage.

## EP 1 755 110 A2

16. Verfahren nach einem vorstehenden Anspruch, bei welchem dem Vorhersage-Ausgangssignal ( $sv(t)$ ) das Audio-Eingangssignal ( $s(t)$ ) zugemischt wird.
- 5 17. Verfahren nach einem vorstehenden Anspruch, bei dem zum Durchführen des Verfahrens ein feldprogrammierbarer Baustein oder eine ASIC (Application-Specified-Integrated-Circuit) entsprechend programmiert wird.
18. Vorrichtung zur Reduktion von Rausch- und Hintergrundsignalen in einem sprachverarbeitenden System, mit
- 10 - einem Audio-Eingang (1) zum Eingeben eines Audio-Eingangssignals ( $s(t)$ ),  
- einem adaptiven Filter (F1) zum Filtern Audio-Eingangssignals ( $s(t)$ ) zum Erzeugen eines Vorhersage-Ausgangssignals ( $sv(t)$ ) mit reduziertem Rauschen,  
- mit einem Speicher zum Speichern einer Vielzahl von Koeffizienten ( $c_i(t)$ ; C1 - C4) für das Filter (F1),  
- wobei das Filter (F1) ausgebildet oder geschaltet ist zur Bildung einer Vielzahl von Vorhersagefehlern ( $sv1 - sv4$ ) und zur Bildung eines Fehlers ( $e$ ) aus der Vielzahl von Vorhersagefehlern ( $sv1 - sv4$ ),
- 15 **dadurch gekennzeichnet, dass**
- eine Koeffizienten-Bereitstellungsanordnung (9) ausgebildet oder geschaltet ist mittels zumindest einem Reduktionsparameter ( $k$ ) die Beträge der Koeffizienten ( $c_i(t)$ ; C1 - C4) fortlaufend zu reduzieren.
- 20 19. Vorrichtung nach Anspruch 18, bei der die Koeffizienten-Bereitstellungsanordnung (9) zum Multiplizieren der Koeffizienten ( $c_i(t)$ ) mit dem Reduktionsparameter ( $k$ ) als einem Faktor  $k$  kleiner 1, insbesondere mit einem Faktor zwischen 0,8 und 1,0 ausgebildet oder geschaltet ist.
- 25 20. Vorrichtung nach Anspruch 18 oder 19, bei welcher einer ersten Filterstufe mit dem Filter als erstem Filter (F1) eine zweite Filterstufe mit einem zweiten Filter (F2) nachgeschaltet ist zum Zuführen des Vorhersage-Ausgangssignals ( $sv(t)$ ) als eine Vorhersage des Audio-Eingangssignals ( $s(t)$ ) mit reduziertem Rauschen als Eingangssignal für den zweiten Filter (F2) zum Erzeugen einer zweiten Vorhersage ( $sv^*(t)$ ).
- 30 21. Vorrichtung nach Anspruch 20 mit einem Addierer (13) zum Addieren einer Summe aus Fehler-Vorhersagen ( $sv^*1 - sv^*4$ ) des zweiten Filters (F2) vom Vorhersage-Ausgangssignal ( $sv(t)$ ) der ersten Filters (F1) zum Erzeugen der Vorhersage ( $sv^*(t)$ ).
- 35 22. Vorrichtung nach einem der Ansprüche 18 bis 21, bei der das zweite Filter (F2) durch ein LMS-Adaptions-Filter zum Durchführen einer Fehlervorhersage ausgebildet oder geschaltet ist.
23. Vorrichtung nach einem der Ansprüche 18 bis 21, bei der das erste Filter (F1) und/oder das zweite Filter (F2) durch ein FIR-Filter zum Durchführen einer Signalvorhersage ausgebildet oder geschaltet ist.
- 40 24. Vorrichtung nach einem der Ansprüche 18 bis 22, welche durch einen feldprogrammierbaren Baustein oder einen ASIC ausgebildet ist.
25. Vorrichtung nach einem der Ansprüche 18 bis 24 mit einer Subtraktionsschaltung (14) zum Abziehen der Werte der Vorhersage ( $sv^*(t)$ ) von Werten des Audio-Eingangssignals ( $s(t)$ ) zum Erzeugen eines rauschreduzierten Audio-Ausgangssignals ( $o^*(t)$ ).
- 45 26. Vorrichtung nach einem der Ansprüche 18 bis 25 mit
- einem Multiplizierer (15; 15\*) zum Wichten des optional zeitlich verzögerten Audio-Eingangssignals ( $s(t)$ ) oder zum Wichten des Vorhersage-Ausgangssignals ( $sv(t)$ ) mit einem Wichtungsfaktor ( $\eta$ ;  $\eta^*$ ) kleiner Eins, insbesondere etwa 0,1 und
- 50 - einem Addierer (7; 14\*) zum Addieren des gewichteten Signals auf das Vorhersage-Ausgangssignal ( $sv(t)$ ) oder auf die Vorhersage ( $sv^*(t)$ ) zum Erzeugen eines rauschreduzierten Audio-Ausgangssignals ( $o(t)$ ;  $o^*(t)$ ).
- 55 27. Vorrichtung nach einem der Ansprüche 18 - 26, die zum Durchführen eines Verfahrens nach einem der Ansprüche 1 - 17 ausgebildet ist.



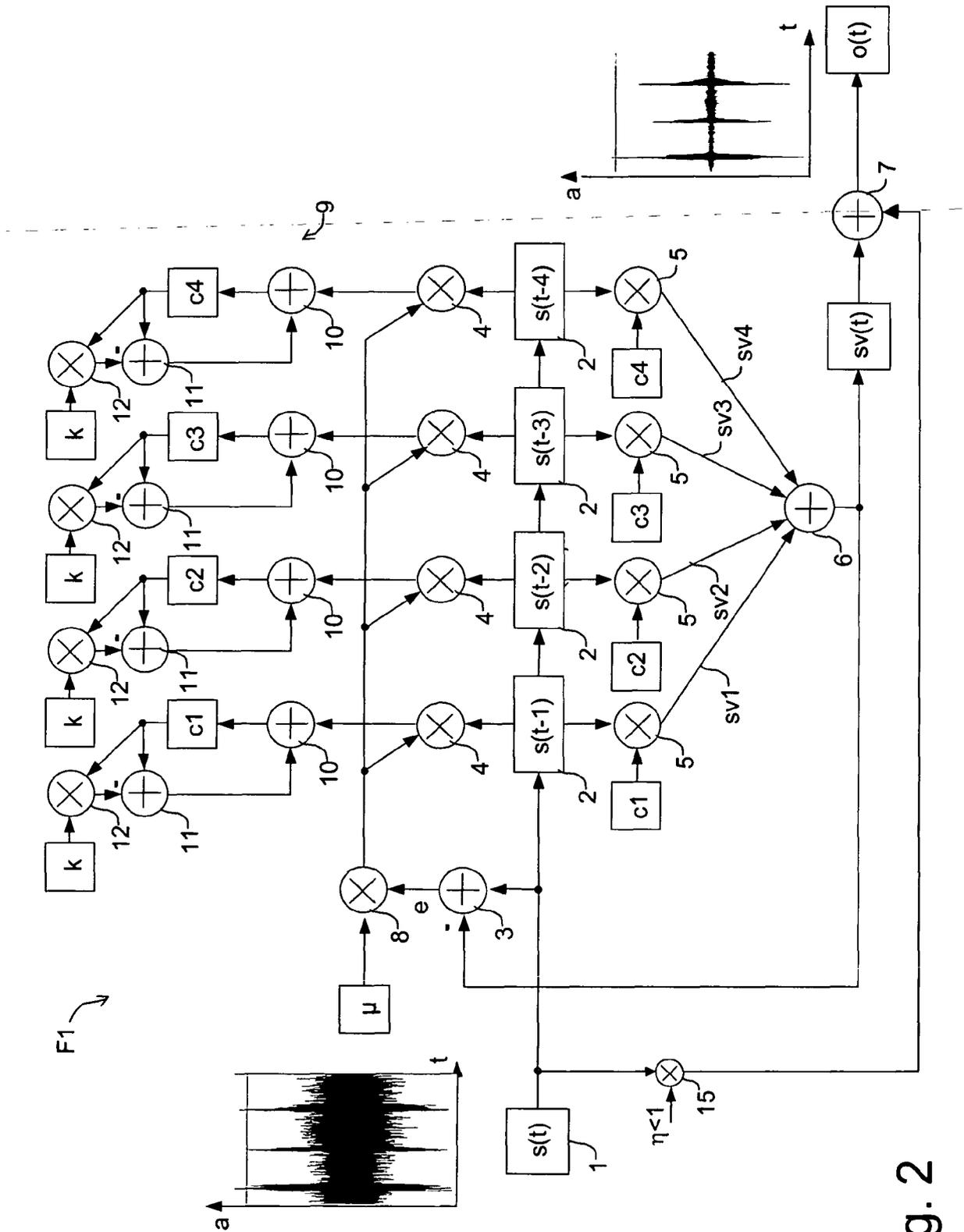


Fig. 2

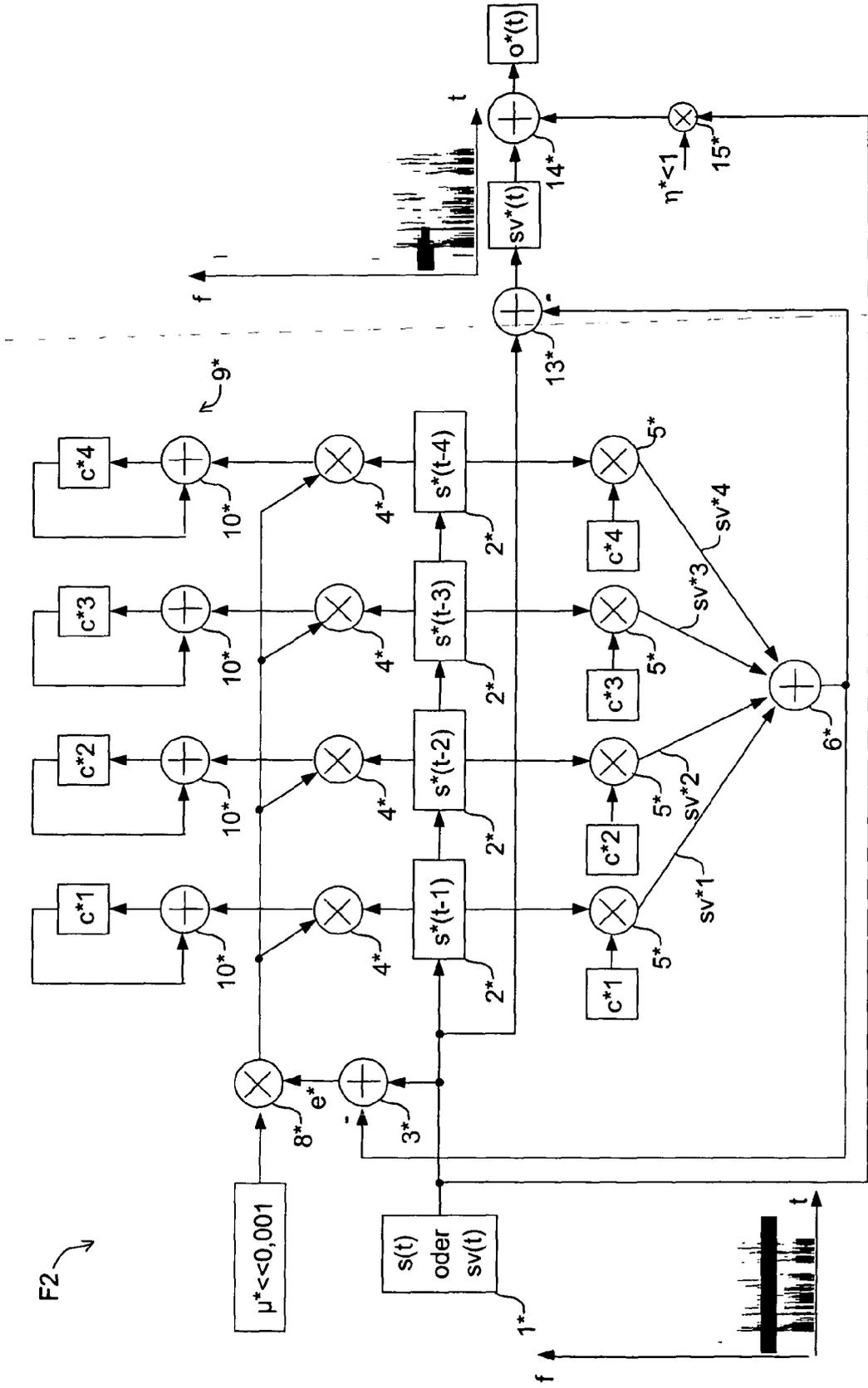


Fig. 3

**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**

- EP 1080465 A [0004] [0032]
- US 6820053 B [0004] [0005] [0032]
- US 5583968 A [0005]
- US 5500903 A [0005]