



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) **DE 695 33 734 T2** 2005.11.03

(12)

## Übersetzung der europäischen Patentschrift

(97) **EP 1 017 042 B1**

(21) Deutsches Aktenzeichen: **695 33 734.3**

(96) Europäisches Aktenzeichen: **00 101 229.3**

(96) Europäischer Anmeldetag: **18.01.1995**

(97) Erstveröffentlichung durch das EPA: **05.07.2000**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **03.11.2004**

(47) Veröffentlichungstag im Patentblatt: **03.11.2005**

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

(30) Unionspriorität:  
**188294**      **28.01.1994**      **US**

(73) Patentinhaber:  
**AT&T Corp., New York, N.Y., US**

(74) Vertreter:  
**derzeit kein Vertreter bestellt**

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

(72) Erfinder:  
**Janiszewski, Thomas John, Andover, Sussex, NJ  
07821, US; Recchione, Michael Charles, Nutley,  
New Jersey 07110, US**

(54) Bezeichnung: **Durch Sprachaktivitätsdetektion gesteuerte Rauschunterdrückung**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

## Beschreibung

**[0001]** Ein Funktelefonsystem umfasst drei wesentliche Elemente: ein zellulares Vermittlungssystem, das als Gateway zum verdrahteten Telefonnetz dient, eine Anzahl von unter der Kontrolle des Vermittlungssystems stehenden Basisstationen, die technische Ausrüstung enthalten, welche zwischen den Signalen, die in dem verdrahteten Telefonnetz verwendet werden, und den Funksignalen, die für drahtlose Kommunikation verwendet werden, übersetzt, und eine Anzahl von Mobiltelefon-Endgeräten, die zwischen den Funksignalen, die für die Kommunikation mit den Basisstationen verwendet werden, und den hörbaren akustischen Signalen, die zur Kommunikation mit den Nutzern verwendet werden (beispielsweise Sprache, Musik usw.), übersetzen.

**[0002]** Die Kommunikation zwischen einer Basisstation und einem Mobiltelefon ist nur möglich, wenn sowohl die Basisstation als auch das Mobiltelefon identische Funkmodulationsschemata, Datenkodierungskonventionen und Steuerungsstrategien verwenden, d. h. beide Einheiten müssen einer Luftschnittstellenspezifikation entsprechen. In den USA wurden bereits eine Reihe von Standards für Luftschnittstellen aufgestellt. Bis vor kurzem basierte die gesamte Funktelefonie in den USA auf dem Advanced Mobile Phone Service (AMPS)-Standard. Dieser Standard schreibt eine analoge Signalkodierung mittels Frequenzmodulation im 800 MHz-Bereich des Funkspektrums vor.

**[0003]** Bei diesem Schema wird jeder Mobiltelefonunterhaltung ein Kommunikationskanal zugewiesen, der für die gesamte Dauer des Telefonats aus zwei 30 kHz-Segmenten dieses Bereichs besteht. Damit sich die Gespräche nicht untereinander stören, dürfen keine zwei Gespräche innerhalb desselben Territoriums gleichzeitig denselben Kanal belegen. Da der Gesamtanteil am Funkspektrum, der der Funktelefonie zugewiesen werden kann, endlich ist, erlegt diese Beschränkung der Anzahl der gleichzeitigen Benutzer eines Funktelefonsystems eine Grenze auf.

**[0004]** Um die Kapazität des Systems zu erhöhen, wurden eine Reihe von Alternativen zum AMPS-Standard eingeführt.

**[0005]** Eine dieser Alternativen ist der Interim Standard-54 (IS-54), der von der Electronic Industries Association und der Telecommunications Industry Association herausgegeben wurde. Dieser Standard verwendet eine digitale Signalkodierung und -modulierung mittels eines zeitüberlappenden Mehrfachzugriffs (Time Division Multiple Access – TDMA). Beim TDMA teilen sich jeweils drei gleichzeitige Gespräche in ein 30 kHz-Segment, und jedes Gespräch darf den Kanal für ein Drittel der Zeit nutzen. Die Zeit ist in Datenblöcke von jeweils 20 ms aufgeteilt, und jeder Datenblock ist weiter in drei Zeitschlitze unterteilt. Jedem Gespräch wird ein Zeitschlitz je Datenblock zugewiesen.

**[0006]** Damit alle Informationen, die 20 ms an Gesprächszeit darstellen, in einem einzigen Zeitschlitz übertragen werden können, werden Sprache und andere Audiosignale mittels eines digitalen Sprachkomprimierungsverfahrens verarbeitet, das als "Vektorensummengesteuerte lineare Prädiktion" (Vector Sum Excited Linear Prediction – VSELP) bezeichnet wird. Jede IS-54-kompatible Basisstation und jedes IS-54-kompatible Mobiltelefon enthält einen VSELP-Kodierer und -Dekodierer. Anstatt eine digitale Repräsentation der Audiowellenform über den Kanal zu übertragen, verwendet der VSELP-Kodierer ein Modell der menschlichen Spracherzeugung, um das digitalisierte Audiosignal auf eine Gruppe von Parametern zu reduzieren, die den Zustand des Spracherzeugungsmechanismus während des Datenpaketes darstellen (beispielsweise die Tonhöhe, die Stimmapparatkonfiguration usw.).

**[0007]** Diese Parameter werden als ein digitaler Bitstrom kodiert und werden dann mit 8 Kilobit pro Sekunde (kbs) über den Kanal zum Empfänger übertragen. Dies ist eine deutlich niedrigere Bitrate, als erforderlich wäre, um die tatsächliche Audiowellenform zu kodieren. Der VSELP-Dekodierer beim Empfänger verwendet dann diese Parameter, um eine neue Schätzung der digitalisierten Audiowellenform zu erzeugen. Die übertragenen digitalen Sprachdaten werden zu digitalen Informationsdatenblöcken von 20 ms organisiert, die jeweils 160 Abtastungen enthalten. Es gibt 159 Bits pro Sprachdatenblock. Das VSELP-Verfahren ist in dem Dokument TR45 Full-Rate Speech Codec Compatibility Standard FN-2972, 1990, herausgegeben von der Electronics Industries Association, (im weiteren als "VSELP-Standard" bezeichnet), detailliert beschrieben.

**[0008]** VSELP verringert deutlich die Anzahl der Bits, die für die Übertragung von Audioinformationen über den Kommunikationskanal benötigt werden. Jedoch stützt es sich bei dieser Verringerung stark auf ein Modell der Spracherzeugung. Folglich werden nicht-sprachliche Töne schlecht wiedergegeben. Beim Innenraum eines fahrenden Kraftfahrzeuges beispielsweise handelt es sich um eine inhärent geräuschvolle Umgebung. Die Eigengeräusche des Kraftfahrzeuges verbinden sich mit äußeren Geräuschen und erzeugen dabei einen

akustischen Hintergrundgeräuschpegel, der weit über dem liegt, den man normalerweise in stillstehenden Umgebungen antrifft. Durch diese Situation ist VSELP gezwungen zu versuchen, über weite Teile der Zeit nicht-sprachliche Informationen sowie Kombinationen aus Sprache und Hintergrundgeräusch zu kodieren.

**[0009]** Wenn zum Kodieren von Sprache in Gegenwart von Hintergrundgeräuschen VSELP verwendet wird, so ergeben sich zwei Probleme: Erstens klingt das Hintergrundgeräusch unnatürlich, unabhängig davon, ob Sprache präsent ist oder nicht, und zweitens wird die Sprache in einer charakteristischen Weise verzerrt. Diese Probleme werden einzeln und zusammen üblicherweise als "Wirbel" bezeichnet.

**[0010]** Zwar wäre es möglich, diese durch den Kodierungs-/Dekodierungsprozess hervorgerufenen Artefakte zu eliminieren, indem man den VSELP-Algorithmus durch einen anderen Sprachkomprimierungsalgorithmus ersetzt, der nicht mit diesen Nachteilen behaftet ist, doch würde diese Strategie eine Änderung der IS-54-Luftschnittstellenspezifikation erfordern. Eine solche Änderung ist aber nicht wünschenswert, weil die Funktelefonie-Serviceanbieter, -Hersteller und -Teilnehmer bereits erhebliche Investitionen in vorhandene Technik getätigt haben. Bei einem Verfahren nach dem Stand der Technik beispielsweise erkennt der Sprachkodierer, wenn keine Sprache anliegt, und kodiert einen speziellen Datenblock, der zum Empfänger übertragen wird. Dieser spezielle Datenblock enthält Komfortgeräuschparameter, die anzeigen, dass der Sprachkodierer Komfortgeräusch erzeugen soll, das dem senderseitigen Hintergrundgeräusch ähnelt. Diese speziellen Datenblöcke werden periodisch durch den Sender während sprachfreier Zeiträume übertragen. Dieser Vorschlag zur Lösung des Wirbel-Problems erfordert eine Änderung am derzeit verwendeten VSELP-Sprachalgorithmus, weil er mit speziellen kodierten Datenblöcken einhergeht, die anzeigen, wann Komfortgeräusch erzeugt werden soll. Er wird sowohl auf der Senderseite als auch auf der Empfängerseite des Kommunikationskanals implementiert und erfordert eine Änderung des derzeitigen Luftschnittstellenspezifikationsstandards, weshalb er keine wünschenswerte Lösung ist.

**[0011]** GB-A-2 256 351 offenbart einen Kommunikationssender/empfänger zum Kommunizieren in Datenblöcken aus kodierten Audiosignalen, umfassend einen Sender und einen Empfänger, einen Sprachkodierer und -dekodierer, einen Sprachaktivitätsdetektor sowie das Erzeugen und Einfügen von Komfortgeräusch während sprachfreier Zeiträume.

**[0012]** WO-A-93 20669 betrifft eine Audioverarbeitungs Vorrichtung mit einer Schaltung zur Geräuschunterdrückung. Die Vorrichtung enthält ein Mikrofon, das Umgebungsgeräuschenergie zu einem Eingangssignal A umwandelt, das durch eine variable hochpassfilternde Schaltung verarbeitet wird. Ein Signal B, das vom Eingangssignal A abgespalten wird, wird durch eine Geräuscherkennungsschaltung verarbeitet, die eine tiefpassfilternde Schaltung und einen Pegeldetektor beinhaltet. Ein Ausgabesignal C von der Geräuscherkennungsschaltung steuert die variable hochpassfilternde Schaltung so, dass sie einen Bereich von Niederfrequenzen des Eingangssignals A proportional zu dem erkannten Geräuschpegel dämpft.

#### Kurzdarstellung der Erfindung

**[0013]** Die Erfindung betrifft ein Verfahren und eine Vorrichtung, wie sie in den unabhängigen Ansprüchen dargelegt sind. Bevorzugte Formen sind in den abhängigen Ansprüchen dargelegt.

**[0014]** Ein Vorteil der vorliegenden Erfindung besteht darin, dass sie das Ausmaß der Artefakte verringern kann, die durch VSELP (oder durch einen anderen Sprachkodierungs/-dekodierungsalgorithmus) hervorgerufen werden, wenn sie in Gegenwart von akustischem Hintergrundgeräusch verwendet wird, ohne dass die Luftschnittstellenspezifikation geändert werden muss.

**[0015]** Es wurde festgestellt, dass eine Kombination aus Signaldämpfung mit Komfortgeräuscheinfügung während sprachfreier Zeiträume und selektiver Hochpassfilterung anhand einer Schätzung der Hintergrundgeräuschenergie eine wirksame Lösung des oben besprochenen Wirbel-Problems darstellt.

**[0016]** In einer Ausführungsform der Erfindung verwendet ein Sprachaktivitätsdetektor eine Energieschätzung zum Erkennen des Vorliegens von Sprache im empfangenen Sprachsignal in einer Geräuschumgebung. Wenn keine Sprache vorliegt, so dämpft das System das Signal und fügt tiefpassgefiltertes weißes Rauschen (d. h. Komfortgeräusch) auf einem geeigneten Pegel ein. Dieses Komfortgeräusch ahmt die typischen spektralen Charakteristika von automobilen oder sonstigen Hintergrundgeräuschen nach. Dadurch wird der Wirbel geglättet und ein natürlicher Klang geschaffen. Wenn der Sprachaktivitätsdetektor feststellt, dass Sprache in dem Signal vorhanden ist, so wird das synthetisierte Sprachsignal ohne Dämpfung verarbeitet.

[0017] Er wurde festgestellt, dass die wahrnehmbar störenden Artefakte, die der Sprachkodierer bei dem Versuch, sowohl Sprache als auch Geräusch zu kodieren, erzeugt, überwiegend im niederfrequenten Bereich auftreten.

[0018] Darum werden – in Abhängigkeit vom Hintergrundgeräuschpegel – neben der sprachaktivitätsgesteuerten Dämpfung und Komfortgeräuscheinfügung eine Reihe Hochpassfilter verwendet. Das Sprachsignal wird dieser Filterung unabhängig davon unterzogen, ob Sprache vorliegt oder nicht. Wenn festgestellt wird, dass der Geräuschpegel unter  $-52$  dB liegt, so wird keine Hochpassfilterung angewendet. Wenn der Geräuschpegel zwischen  $-40$  dB und  $-52$  dB liegt, so wird das synthetisierte Sprachsignal einer Hochpassfilterung mit einer Grenzfrequenz von  $200$  Hz unterzogen. Wenn der Geräuschpegel über  $-40$  dB liegt, so findet eine Hochpassfilterung mit einer Grenzfrequenz von  $350$  Hz statt. Das Ergebnis dieser Hochpassfilter ist ein verringertes Hintergrundgeräusch mit minimaler Auswirkung auf die Sprachqualität.

[0019] Die hier beschriebene Erfindung wird beim Empfänger implementiert (entweder in der Basisstation, dem Mobiltelefon oder in beiden) und kann damit realisiert werden, ohne das derzeitige standardmäßige Sprachkodierungs/-dekodierungsprotokoll zu ändern.

#### KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0020] [Fig. 1](#) ist ein Blockschaubild eines digitalen Funkempfangssystems, das die vorliegende Erfindung beinhaltet.

[0021] [Fig. 2](#) ist ein Blockschaubild eines durch Sprachaktivitätserkennung gesteuerten Geräuschminderers, der die vorliegende Erfindung beinhaltet.

[0022] [Fig. 3](#) ist eine Wellenform, welche die akustische Gesamtenergie eines empfangenen Signals darstellt.

[0023] [Fig. 4](#) ist ein Blockschaubild eines Hochpassfiltertreibers.

[0024] [Fig. 5](#) ist ein Ablaufdiagramm der Funktion des Sprachaktivitätsdetektors.

[0025] [Fig. 6](#) zeigt ein Blockschaubild einer Mikroprozessorverkörperung der vorliegenden Erfindung.

#### DETAILLIERTE BESCHREIBUNG

[0026] In [Fig. 1](#) ist ein digitales Funkempfangssystem **10** gezeigt, das die vorliegende Erfindung beinhaltet. Ein Demodulator **20** empfängt übertragene Wellenformen, die kodierten Sprachsignalen entsprechen, und verarbeitet die empfangenen Wellenformen zu einem digitalen Signal **d**. Dieses digitale Signal **d** wird einem Kanaldekodierer **30** zugeleitet, der das Signal **d** so verarbeitet, dass Kanalfehler gemindert werden. Das resultierende Signal, das vom Kanaldekodierer **30** erzeugt wird, ist ein kodierter Sprachbitstrom **b**, der nach dem VSELP-Standard, der oben im Zusammenhang mit dem allgemeinen Stand der Technik beschrieben wurde, zu digitalen Informationsdatenblöcken organisiert wird. Dieser kodierte Sprachbitstrom **b** wird einem Sprachdekodierer **40** zugeführt, der den kodierten Sprachbitstrom **b** zu einem dekodierten Sprachbitstrom **s** verarbeitet. Dieser Sprachdekodierer **40** ist dafür konfiguriert, Sprache zu dekodieren, die mit der VSELP-Technik kodiert wurde. Dieser dekodierte Sprachbitstrom **s** wird einem über Sprachaktivitätserkennung gesteuerten Geräuschminderer (Voice Activity Detection Driven Noise Remediator – VADDNR) **50** zugeführt, um alle Hintergrund-"Wirbel" zu entfernen, die während sprachfreien Zeiträumen in dem Signal vorhanden sind. In einer Ausführungsform empfängt der VADDNR **50** auch einen Teil des kodierten Sprachbitstroms **b** vom Kanaldekodierer **30** über die Signalleitung **35**. Der VADDNR **50** verwendet den VSELP-kodierten Datenblockenergiewert  $r_0$ , der ein Teil des kodierten Bitstroms **b** ist, wie weiter unten eingehender besprochen werden wird.

[0027] Der VADDNR **50** erzeugt eine verarbeitete dekodierte Sprachbitstromausgabe **s''**. Die Ausgabe vom VADDNR **50** kann dann einem Digital-Analog-Wandler **60** zugeführt werden, das das digitale Signal **s''** in eine analoge Wellenform umwandelt. Diese analoge Wellenform kann dann zu einem Zielsystem, wie beispielsweise einem Telefonnetz, gesandt werden. Alternativ kann die Ausgabe vom VADDNR **50** auch einer anderen Vorrichtung zugeführt werden, welche die Ausgabe vom VADDNR in ein anderes digitales Dateiformat, das von einem Zielsystem verwendet wird, umwandelt.

[0028] Der VADDNR **50** ist in [Fig. 2](#) detaillierter dargestellt.

[0029] Der VADDNR empfängt den VSELP-kodierten Datenblockenergiwert  $r_0$  vom dem kodierten Sprachbitstrom  $b$  über die Signalleitung **35**, wie in [Fig. 1](#) gezeigt. Dieser Energiwert  $r_0$  repräsentiert die durchschnittliche Signalstärke in der Eingabesprache über den 20 ms-Datenblockintervall hinweg. Es gibt 32 mögliche Werte für  $r_0$ , und zwar 0 bis 31.  $r_0 = 0$  repräsentiert eine Datenblockenergie von 0. Die übrigen Werte für  $r_0$  reichen von einem Minimum von  $-64$  dB, was  $r_0 = 1$  entspricht, bis zu einem Maximum von  $-4$  dB, was  $r_0 = 31$  entspricht. Die Schrittgröße zwischen  $r_0$ -Werten beträgt 2 dB. Der Datenblockenergiwert  $r_0$  wird in VSELP-Standard, S. 16, eingehender beschrieben. Der kodierte Datenblockenergiwert  $r_0$  wird einem Energieschätzer **210** zugeführt, der die durchschnittliche Datenblockenergie bestimmt.

[0030] Der Energieschätzer **210** erzeugt ein Signal  $e[m]$  der durchschnittlichen Datenblockenergie, das die durchschnittliche Datenblockenergie repräsentiert, die während eines Datenblocks  $m$  berechnet wurde, wobei  $m$  = ein Datenblockindex, der den aktuellen digitalen Informationsdatenblock repräsentiert.

[0031]  $e[m]$  ist definiert als:

$$e[m] = \begin{cases} E_{init} & \text{für } m = 0 \\ \alpha * r_0[m] + (1-\alpha) * e[m-1] & \text{für } m > 0 \end{cases}$$

[0032] Die durchschnittliche Datenblockenergie wird anfangs auf eine erste Energieschätzung  $E_{init}$  eingestellt.

[0033]  $E_{init}$  wird auf einen Wert größer als 31 eingestellt, wobei es sich um den größtmöglichen Wert für  $r_0$  handelt. Beispielsweise könnte  $E_{init}$  auf einen Wert von 32 eingestellt werden. Nach der Initialisierung wird die durchschnittliche Datenblockenergie  $e[m]$  mittels der Gleichung  $e[m] = \alpha * r_0[m] + (1 - \alpha) * e[m - 1]$  berechnet, wobei  $\alpha$  = eine Glättungskonstante mit  $0 \leq \alpha \leq 1$ .  $\alpha$  sollte so gewählt werden, dass sich eine akzeptable Datenblockdurchschnittsbildung ergibt. Wie haben festgestellt, dass ein Wert von  $\alpha = 0,25$  optimal ist, weil er eine effektive Datenblockdurchschnittsbildung über sieben Blöcke aus digitalen Informationen (140 ms) erbringt. Es können für  $\alpha$  auch andere Werte gewählt werden, wobei der Wert vorzugsweise im Bereich von  $0,25 \pm 0,2$  liegt.

[0034] Wie oben besprochen, und wie in [Fig. 1](#) zu sehen, empfängt der VADDNR **50** den VSELP-kodierten Datenblockenergiwert  $r_0$  von dem kodierten Sprachbitstromsignal  $b$ , bevor das Signal  $b$  durch den Sprachdekodierer **40** dekodiert wird. Alternativ kann dieser Datenblockenergiwert  $r_0$  auch durch den VADDNR **50** selbst aus dem dekodierten Sprachbitstromsignal  $s$  berechnet werden, das vom Sprachdekodierer **40** kommt. In einer Ausführungsform, wo der Datenblockenergiwert  $r_0$  durch den VADDNR **50** berechnet wird, braucht dem VADDNR **50** kein Teil des kodierten Sprachbitstroms  $b$  zugeführt werden, und die Signalleitung **35**, die in [Fig. 1](#) zu sehen ist, wäre nicht vorhanden. Stattdessen würde der VADDNR **50** nur den dekodierten Sprachbitstrom  $s$  verarbeiten, und der Datenblockenergiwert  $r_0$  würde so berechnet werden, wie es in VSELP-Standard, S. 16–17, beschrieben ist. Wenn aber  $r_0$  dem VADDNR **50** von dem kodierten Bitstrom  $b$  über die Signalleitung **35** zugeführt wird, so kann der VADDNR den dekodierten Sprachbitstrom  $s$  schneller berechnen, weil er  $r_0$  nicht berechnen muss.

[0035] Das vom Energieschätzer **210** erzeugte Signal  $e[m]$  der durchschnittlichen Datenblockenergie repräsentiert die durchschnittliche akustische Gesamtenergie, die in dem empfangenen Sprachsignal vorhanden ist. Diese akustische Gesamtenergie kann sowohl aus Sprache als auch Geräusch zusammengesetzt sein. Als Beispiel zeigt [Fig. 3](#) eine Wellenform, welche die akustische Gesamtenergie eines typischen empfangenen Signals **310** im zeitlichen Verlauf  $T$  darstellt. In einer sich bewegenden Umgebung wird es in der Regel immer einen bestimmten Umgebungshintergrundgeräuschpegel geben. Der Energiepegel dieses Geräuschs ist in [Fig. 3](#) als  $e_1$  dargestellt. Wenn im Signal **310** Sprache vorhanden ist, so repräsentiert der akustische Energiepegel sowohl Sprache als auch Geräusch. Das ist in [Fig. 3](#) in dem Bereich gezeigt, wo die Energie  $> e_2$ . Während des Zeitintervalls  $t_1$  ist im Signal **310** keine Sprache vorhanden, und die akustische Energie während dieses Zeitintervalls  $t_1$  repräsentiert lediglich Umgebungshintergrundgeräusch. Während des Zeitintervalls  $t_2$  ist im Signal **310** Sprache vorhanden, und die akustische Energie während dieses Zeitintervalls  $t_2$  repräsentiert Umgebungshintergrundgeräusch und Sprache.

[0036] In [Fig. 2](#) ist gezeigt, wie das vom Energieschätzer **210** erzeugte Ausgabesignal  $e[m]$  einem Geräuschschätzer **220** zugeführt wird, der den durchschnittlichen Hintergrundgeräuschpegel in dem dekodierten Sprachbitstrom  $s$  bestimmt. Der Geräuschschätzer **220** erzeugt ein Signal  $N[m]$ , das einen Geräuschschätzwert repräsentiert, wobei:

**Ninit**für **m = 0****N[m] = N[m-1]**für **e[m] > N[m-1] + Nschwelle** **$\beta * e[m] + (1-\beta) * N[m-1]$**  ... andernfalls

[0037] N[m] wird zunächst auf den Anfangswert Ninit eingestellt, bei dem es sich um eine anfängliche Geräuschschätzung handelt. Im Verlauf der weiteren Verarbeitung erhöht oder verringert sich der Wert N[m] je nach dem tatsächlichen Hintergrundgeräusch, das in dem dekodierten Sprachbitstrom s vorhanden ist. Ninit wird auf einen Pegel eingestellt, der an der Grenze zwischen moderatem und starkem Hintergrundgeräusch liegt. Das Initialisieren von N[m] auf diesen Pegel ermöglicht es, dass N[m] sich rasch in beiden Richtungen anpassen kann, so wie es durch das tatsächliche Hintergrundgeräusch bestimmt wird. Wir haben festgestellt, dass es in einer sich bewegendenden Umgebung bevorzugt ist, Ninit auf einen r0-Wert von 13 einzustellen.

[0038] Die Sprachkomponente der Signalenergie sollte nicht in die Berechnung des durchschnittlichen Hintergrundgeräuschpegels aufgenommen werden. In [Fig. 3](#) sollte beispielsweise der Energiepegel, der während des Zeitintervalls  $t_1$  im Signal **310** anliegt, in die Berechnung der Geräuschschätzung N[m] aufgenommen werden, aber der Energiepegel, der während des Zeitintervalls  $t_2$  im Signal **310** anliegt, in nicht aufgenommen werden, weil die Energie während des Zeitintervalls  $t_2$  sowohl Hintergrundgeräusch als auch Sprache repräsentiert.

[0039] Darum sollte jede durchschnittliche Datenblockenergie e[m], die vom Energieschätzer **210** stammt und sowohl Sprache als auch Geräusch repräsentiert, aus der Berechnung der Geräuschschätzung N[m] ausgeklammert werden, damit die Geräuschschätzung N[m] nicht verzerrt wird. Um Werte durchschnittlicher Datenblockenergie e[m], die sowohl Sprache als auch Geräusch repräsentieren, auszuklammern, wird eine obere Geräuschausblendungsschwelle "Nschwelle" verwendet. Wenn also, wie oben dargelegt,  $e[m] > N[m - 1] + Nschwelle$ , dann ist  $N[m] = N[m - 1]$ . Anders ausgedrückt: Wenn die durchschnittliche Datenblockenergie e[m] des aktuellen Datenblocks um einen Betrag, der mindestens so groß ist wie Nschwelle, größer ist die Geräuschschätzung N[M - 1] des vorangegangenen Datenblocks, d. h. wenn Sprache vorliegt, so wird N[m] nicht gegenüber der Berechnung der vorangegangenen Datenblocks geändert. Wenn also die Datenblockenergie über einen kurzen Zeitraum hinweg stark ansteigt, so wird davon ausgegangen, dass dieser Anstieg auf das Vorhandensein von Sprache zurückzuführen ist, und die Energie wird nicht in die Geräuschschätzung aufgenommen. Wir haben festgestellt, dass es optimal ist, Nschwelle auf das Äquivalent eines Wertes der Datenblockenergie r0 von 2,5 einzustellen. Dies beschränkt den Arbeitsbereich des Geräuschschätzungsalgorithmus' auf Bedingungen, die besser sind als ein Audiosignal-Geräusch-Verhältnis von 5 dB, weil r0 in Einheiten von 2 dB skaliert wird. Nschwelle könnte auf einen beliebigen Punkt im Bereich von 2 bis 4 für eine akzeptable Leistung des Geräuschschätzers **220** eingestellt werden.

[0040] Wenn die Datenblockenergie keinen starken Anstieg über einen kurzen Zeitraum hinweg erfährt, so wird die Geräuschschätzung anhand der Gleichung  $N[m] = \beta * e[m] + (1 - \beta) * N[m - 1]$  bestimmt, wobei  $\beta$  eine Glättungskonstante ist, die so eingestellt werden sollte, dass eine akzeptable Datenblockdurchschnittsbildung gewährleistet ist. Es wurde festgestellt, dass ein Wert von 0,05 für  $\beta$ , was eine Datenblockdurchschnittsbildung über 25 Datenblöcke (500 ms) ergibt, bevorzugt ist. Der Wert von  $\beta$  sollte allgemein auf den Bereich von  $0,025 \leq \beta \leq 0,1$  eingestellt werden.

[0041] Der durch den Geräuschschätzer **220** berechnete Geräuschschätzungswert N[m] wird einem Hochpassfiltertreiber **260** zugeführt, der auf der Basis des dekodierten Bitstromsignals s arbeitet, das vom Sprachdekodierer **40** stammt. Wie oben besprochen, enthält jeder digitale Informationsdatenblock **160** Abtastungen von Sprachdaten.

[0042] Der Hochpassfiltertreiber **260** arbeitet auf der Basis jeder dieser Abtastungen s[i], wobei i = ein Abtastindex. Der Hochpassfiltertreiber **260** ist in [Fig. 4](#) noch eingehender gezeigt. Der durch den Geräuschschätzer **220** erzeugte Geräuschschätzungswert N[m] wird dem Logikblock **410** zugeführt, der Logikschaltungen enthält, um zu bestimmen, welche einer Gruppe von Hochpassfiltern dazu verwendet wird, jede Abtastung s[i] des dekodierten Sprachbitstroms s zu filtern. Es gibt zwei Hochpassfilter: **430** und **440**. Filter **430** hat eine Grenzfrequenz bei 200 Hz, und Filter **440** hat eine Grenzfrequenz bei 350 Hz. Es wurde festgestellt, dass diese Grenzfrequenzen zu optimalen Ergebnissen führen, obgleich auch andere Werte gemäß der vorliegenden Erfindung verwendet werden können. Die Differenz zwischen den Grenzfrequenzen der Filter sollte vorzugsweise wenigstens 100 Hz betragen. Um zu bestimmen, welches Filter verwendet werden sollte, vergleicht der Logikblock **410** des Hochpassfiltertreibers **260** den Geräuschschätzungswert N[m] mit zwei Schwellen. Die erste

Schwelle wird auf einen Wert eingestellt, der einem Datenblockenergiewert  $r_0 = 7$  entspricht (was  $-52$  dB entspricht), und die zweite Schwelle wird auf einen Wert eingestellt, der einem Datenblockenergiewert  $r_0 = 13$  entspricht (was  $-40$  dB entspricht). Wenn die Geräuschschätzung  $N[m]$  niedriger ist als  $r_0 = 7$ , so erfolgt keine Hochpassfilterung.

**[0043]** Wenn der Geräuschschätzungswert  $N[m]$  mindestens  $r_0 = 7$  und weniger als  $r_0 = 13$  beträgt, so wird der 200 Hz-Hochpassfilter **430** angewendet. Wenn der Geräuschschätzungswert  $N[m]$  mindestens  $r_0 = 13$  beträgt, so wird der 350 Hz-Hochpassfilter **440** angewendet. Die Logik für das Bestimmen der anzuwendenden Hochpassfilterung lässt sich folgendermaßen zusammenfassen:

	alle Pässe	für $N[m] < 7$
Filter =	Hochpass bei 200 Hz	für $7 \leq N[m] < 13$
	Hochpass bei 350 Hz	für $N[m] \geq 13$

**[0044]** In [Fig. 4](#) ist gezeigt, wie die Logik durch den Logikblock **410** ausgeführt wird. Der Logikblock **410** bestimmt anhand der obigen Regeln, welches Filter angewendet werden soll, und sendet ein Steuersignal  $c[m]$  an zwei Kreuzschienenschalter **420**, **450**. Ein Steuersignal, das einem Wert von 0 entspricht, zeigt an, dass keine Hochpassfilterung anzuwenden ist. Ein Steuersignal, das einem Wert von 1 entspricht, zeigt an, dass das 200 Hz-Hochpassfilter anzuwenden ist. Ein Steuersignal, das einem Wert von 2 entspricht, zeigt an, dass das 350 Hz-Hochpassfilter anzuwenden ist.

**[0045]** Das Signal  $s[i]$  wird vom Sprachdekodierer **40** an den Kreuzschienenschalter **420** gesendet. Der Kreuzschienenschalter **420** leitet das Signal  $s[i]$  zur entsprechenden Signalleitung **421**, **422**, **423**, um die richtige Filterung auszuwählen. Ein Steuersignal von 0 leitet das Signal  $s[i]$  zur Signalleitung **421**. Die Signalleitung **421** überträgt das Signal  $s[i]$  zum Kreuzschienenschalter **450**, ohne dass eine Hochpassfilterung stattfindet. Ein Steuersignal von 1 leitet das Signal  $s[i]$  zur Signalleitung **422**, die mit dem Hochpassfilter **430** verbunden ist. Nachdem das Signal  $s[i]$  durch das Hochpassfilter **430** gefiltert wurde, wird es über die Signalleitung **424** zum Kreuzschienenschalter **450** übertragen. Ein Steuersignal von 2 leitet das Signal  $s[i]$  zur Signalleitung **423**, die mit dem Hochpassfilter **440** verbunden ist.

**[0046]** Nachdem das Signal  $s[i]$  durch das Hochpassfilter **440** gefiltert wurde, wird es über die Signalleitung **425** zum Kreuzschienenschalter **450** übertragen. Das Steuersignal  $c[m]$  wird ebenfalls zum Kreuzschienenschalter **450** übertragen. Anhand des Steuersignals  $c[m]$  übermittelt der Kreuzschienenschalter **450** eines der Signale von der Signalleitung **421**, **424**, **425** zum Sprachdämpfer **270**.

**[0047]** Dieses durch den Hochpassfiltertreiber **260** erzeugte Signal wird als  $s'[i]$  gekennzeichnet. Der Fachmann erkennt, dass jede beliebige Anzahl von Hochpassfiltern oder ein einziges Hochpassfilter mit einer kontinuierlich einstellbaren Grenzfrequenz im Hochpassfiltertreiber **260** verwendet werden könnte, um den dekodierten Bitstrom  $s$  zu filtern. Die Verwendung einer größeren Anzahl von Hochpassfiltern oder eines einzelnen Hochpassfilters mit einer kontinuierlich einstellbaren Grenzfrequenz würde zum Ergebnis haben, dass die Übergänge zwischen den Filterauswahlen weniger wahrnehmbar wären.

**[0048]** In [Fig. 2](#) ist zu sehen, wie das durch den Hochpassfiltertreiber **260** erzeugte Signal  $s'[i]$  einem Sprachdämpfer/Komfortgeräuscheinkoppler **270** zugeführt wird. Der Sprachdämpfer/Komfortgeräuscheinkoppler **270** verarbeitet das Signal  $s'[i]$  zu dem verarbeiteten dekodierten Sprachbitstrom-Ausgabesignal  $s''[i]$ . Der Sprachdämpfer/Komfortgeräuscheinkoppler **270** empfängt außerdem ein Eingabesignal  $n[i]$  von einem Geräuschformungsgenerator **250** und ein Eingabesignal  $dämpf[m]$  von einem Dämpfungsberechner **240**. Die Funktionsweise des Sprachdämpfers/Komfortgeräuscheinkopplers **270** wird weiter unten näher besprochen, nachdem besprochen wurde, wie seine Eingaben  $n[i]$  und  $dämpf[m]$  berechnet werden.

**[0049]** Die durch den Geräuschschätzer **220** erzeugte Geräuschschätzung  $N[m]$  und die durch den Energieschätzer **210** erzeugte durchschnittliche Datenblockenergie  $e[m]$  werden dem Sprachaktivitätsdetektor **230** zugeführt. Der Sprachaktivitätsdetektor **230** bestimmt, ob in dem aktuellen Datenblock des Sprachsignals Sprache vorhanden ist oder nicht, und erzeugt ein Spracherkennungssignal  $v[m]$ , das anzeigt, ob Sprache vorhanden ist oder nicht.

**[0050]** Ein Wert von 0 für  $v[m]$  zeigt an, dass in dem aktuellen Datenblock des Sprachsignals keine Sprachaktivität erkannt wird. Ein Wert von 1 für  $v[m]$  zeigt an, dass in dem aktuellen Datenblock des Sprachsignals Sprachaktivität erkannt wird. Die Funktionsweise des Sprachaktivitätsdetektors **230** wird in Verbindung mit

dem Ablaufdiagramm von [Fig. 5](#) beschrieben. In Schritt **505** bestimmt der Sprachaktivitätsdetektor **230**, ob  $e[m] < N[m] + T_{\text{detekt}}$ , wobei  $T_{\text{detekt}}$  eine niedrigere Geräuscherkennungsschwelle ist und in ihrer Funktion dem oben in Verbindung mit [Fig. 3](#) besprochenen  $N_{\text{schwelle}}$ -Wert ähnelt. Es wird von der Annahme ausgegangen, dass Sprache nur dann vorliegt, wenn die durchschnittliche Datenblockenergie  $e[m]$  um einen Wert  $-T_{\text{detekt}}$  größer ist als der Geräuschschätzungswert  $N[m]$ .  $T_{\text{detekt}}$  wird vorzugsweise auf einen  $r_0$ -Wert von 2,5 eingestellt, was bedeutet, dass Sprache nur dann vorliegt, wenn die durchschnittliche Datenblockenergie  $e[m]$  um 5 dB größer ist als der Geräuschschätzungswert  $N[m]$ . Es können auch andere Werte verwendet werden. Der Wert von  $T_{\text{detekt}}$  sollte allgemein im Bereich von  $2,5 \pm 0,5$  liegen.

**[0051]** Um zu verhindern, dass der Sprachaktivitätsdetektor **230** während gesprochener Worte "keine Sprachaktivität" signalisiert, wird ein Zähler  $N_{\text{zähl}}$  für nicht-erkannte Datenblöcke verwendet.  $N_{\text{zähl}}$  wird auf Null initialisiert und wird so eingestellt, dass er bis zu einer Schwelle zählt.  $N_{\text{zählschwelle}}$  repräsentiert die Anzahl der sprachaktivitätsfreien Datenblöcke, die vorliegen müssen, bevor der Sprachaktivitätsdetektor **230** signalisiert, dass keine Sprachaktivität vorhanden ist.

**[0052]**  $N_{\text{zählschwelle}}$  kann auf einen Wert von sechs eingestellt werden. Das heißt, erst wenn sechs Datenblöcke lang (120 ms) keine Sprache erkannt wird, signalisiert der Sprachaktivitätsdetektor **230** "keine Sprache". Wenden wir uns wieder [Fig. 5](#) zu. Wenn Schritt **505** bestimmt, dass  $e[m] < N[m] + T_{\text{detekt}}$ , d. h. wenn die durchschnittliche Energie  $e[m]$  niedriger ist als jene, für die bestimmt wurde, dass Sprache vorhanden ist, so wird  $N_{\text{zähl}}$  in Schritt **510** um den Zählwert 1 erhöht. Wenn Schritt **515** bestimmt, dass  $N_{\text{zähl}} \geq N_{\text{zählschwelle}}$ , d. h. dass sechs Datenblöcke vorlagen, in denen keine Sprache erkannt wurde, so wird  $v[m]$  in Schritt **530** auf 0 gesetzt, um "keine Sprache" für den aktuellen Datenblock anzuzeigen. Wenn Schritt **515** bestimmt, dass  $N_{\text{zähl}} < N_{\text{zählschwelle}}$ , d. h. dass noch keine sechs Datenblöcke vorlagen, in denen keine Sprache erkannt wurde, so wird  $v[m]$  in Schritt **520** auf 1 gesetzt, um anzuzeigen, dass im aktuellen Datenblock Sprache vorhanden ist. Wenn Schritt **505** bestimmt, dass  $e[m] \geq N[m] + T_{\text{detekt}}$ , d. h. wenn die durchschnittliche Energie  $e[m]$  mindestens so groß ist wie jene, für die bestimmt wurde, dass Sprache vorhanden ist, so wird  $N_{\text{zähl}}$  in Schritt **525** auf Null gesetzt, und  $v[m]$  wird in Schritt **520** auf 1 gesetzt, um anzuzeigen, dass im aktuellen Datenblock Sprache vorhanden ist.

**[0053]** Das durch den Sprachaktivitätsdetektor **230** erzeugte Spracherkennungssignal  $v[m]$  wird dem Dämpfungsberechner **240** zugeführt, der ein Dämpfungssignal  $d_{\text{ämpf}}[m]$  erzeugt, welches das Maß an Dämpfung des aktuellen Datenblocks repräsentiert. Das Dämpfungssignal  $d_{\text{ämpf}}[m]$  wird mit jedem Datenblock aktualisiert, und sein Wert hängt zum Teil davon ab, ob der Sprachaktivitätsdetektor **230** Sprache erkannt hat oder nicht. Das Signal  $d_{\text{ämpf}}[m]$  repräsentiert einen Wert zwischen 0 und 1. Je näher der Wert bei 1 liegt, desto geringer ist die Dämpfung des Signals, und je näher der Wert bei 0 liegt, desto stärker ist die Dämpfung des Signals. Die maximale Dämpfung, die anzuwenden ist, ist als  $\text{maxd}_{\text{ämpf}}$  definiert, und es wurde festgestellt, dass der optimale Wert für  $\text{maxd}_{\text{ämpf}}$  bei 0,65 (d. h. -3,7 dB) liegt. Es können aber auch andere Werte für  $\text{maxd}_{\text{ämpf}}$  verwendet werden, wobei der Wert allgemein im Bereich von 0,3 bis 0,8 liegt. Der Faktor, um den die Dämpfung der Sprache erhöht wird, ist als  $\text{d}_{\text{ämpfrate}}$  definiert, und es wurde festgestellt, dass der bevorzugte Wert für  $\text{d}_{\text{ämpfrate}}$  bei 0,98 liegt. Es können aber auch andere Werte für  $\text{d}_{\text{ämpfrate}}$  verwendet werden, wobei der Wert allgemein im Bereich von  $0,95 \pm 0,04$  liegt.

**[0054]** In diesem Abschnitt beschreiben wir die Berechnung des Dämpfungssignals  $d_{\text{ämpf}}[m]$ . Die Verwendung von  $d_{\text{ämpf}}[m]$  bei der Dämpfung des Signals  $s'[i]$  wird während der weiter unten folgenden Besprechung in Verbindung mit dem Sprachdämpfer/Komfortgeräuschein Koppler **270** deutlich. Das Dämpfungssignal  $d_{\text{ämpf}}[m]$  wird folgendermaßen berechnet. Das Dämpfungssignal  $d_{\text{ämpf}}[m]$  wird zu Beginn auf 1 gesetzt. Nach dieser Initialisierung wird  $d_{\text{ämpf}}[m]$  in Abhängigkeit davon berechnet, ob Sprache vorliegt, was durch den Sprachaktivitätsdetektor **230** bestimmt wird, und davon, ob die Dämpfung die maximale Dämpfung erreicht hat, die durch  $\text{maxd}_{\text{ämpf}}$  definiert wird. Wenn  $v[m] = 1$ , d. h. wenn Sprache erkannt wird, so wird  $d_{\text{ämpf}}[m]$  auf 1 gesetzt. Wenn  $v[m] = 0$ , d. h. wenn keine Sprache erkannt wird, und wenn der Dämpfungsfaktor, der auf die Dämpfung des vorangegangenen Datenblocks angewendet wurde ( $\text{d}_{\text{ämpfrate}} * d_{\text{ämpf}}[m - 1]$ ), größer ist als die maximale Dämpfung, so wird die Dämpfung des aktuellen Datenblocks durch Anwenden des Dämpfungsfaktors auf die Dämpfung des vorangegangenen Datenblocks berechnet. Wenn  $v[m] = 0$ , d. h. wenn keine Sprache erkannt wird, und wenn der Dämpfungsfaktor, der auf die Dämpfung des vorangegangenen Datenblocks angewendet wurde, maximal so groß ist wie die maximale Dämpfung, so wird die Dämpfung des aktuellen Datenblocks auf die maximale Dämpfung eingestellt. Diese Berechnung der Dämpfung des aktuellen Datenblocks wird folgendermaßen zusammengefasst:



1,0

dämpf[m] = dämpfrate\*dämpf[m - 1]

maxdämpf

für m = 0 oder v[m] = 1

für dämpfrate\*dämpf[m - 1] &gt; maxdämpf und v[m] = 0

für dämpfrate\*dämpf[m - 1] ≤ maxdämpf und v[m] = 0

**[0055]** Wenn also vom Sprachaktivitätsdetektor **230** keine Sprache erkannt wird, so wird das Dämpfungssignal dämpf[m] von 1 auf 0,65 (maxdämpf) um einen Konstantfaktor von 0,98 verringert. Das vom Dämpfungsberechner **240** erzeugte Signal dämpf[m] der Dämpfung des aktuellen Datenblocks wird dem Sprachdämpfer/Komfortgeräuscheinkoppler **270** zugeführt.

**[0056]** Der Sprachdämpfer/Komfortgeräuscheinkoppler **270** empfängt außerdem das Signal n[i], das tiefpassgefiltertes weißes Rauschen repräsentiert, vom Geräuschformungsgenerator **250**. Dieses tiefpassgefilterte weiße Rauschen wird auch als Komfortgeräusch bezeichnet. Der Geräuschformungsgenerator **250** empfängt die Geräuschschätzung N[m] vom Geräuschschätzer **220** und erzeugt das Signal n[i], welches das geformte Geräusch folgendermaßen repräsentiert:

$$n[i] = \varepsilon * wn[i] + (1 - \varepsilon) * n[i - 1],$$

wobei

$$wn[i] = \delta * dB21in(N[m]) * ran[i]$$

wobei i = der Abtastindex, wie oben besprochen. Somit wird n[i] für jede Abtastung in dem aktuellen Datenblock erzeugt. Die Funktion dB21in mappt die Geräuschschätzung N[m] von einem dB-Wert zu einem linearen Wert. Der Skalierungsfaktor δ wird auf einen Wert von 1,7 eingestellt, und der Filterkoeffizient ε wird auf einen Wert von 0,1 eingestellt. Die Funktion ran[i] erzeugt eine Zufallszahl zwischen -1,0 und 1,0. Somit wird das Geräusch mittels der Geräuschschätzung N[m] skaliert und anschließend mit einem Tiefpassfilter gefiltert. Die oben genannten Werte für den Skalierungsfaktor δ und den Filterkoeffizienten ε wurden als optimal befunden. Es können aber auch andere Werte verwendet werden, wobei der Wert für δ allgemein im Bereich von 1,5 bis 2,0 liegt und der Wert für ε allgemein im Bereich von 0,05 bis 0,15 liegt.

**[0057]** Das vom Geräuschformungsgenerator **220** erzeugte tiefpassgefilterte weiße Rauschen n[i] und die vom Dämpfungsberechner **240** erzeugte Dämpfung dämpf[m] des aktuellen Datenblocks werden dem Sprachdämpfer/Komfortgeräuscheinkoppler **270** zugeführt. Der Sprachdämpfer empfängt das hochpassgefilterte Signal s'[i] vom Hochpassfiltertreiber **260** und erzeugt den verarbeiteten dekodierten Sprachbitstrom s" gemäß folgender Gleichung:

$$s''[i] = dämpf[m] * s'[i] + (1 - dämpf[m]) * n[i],$$

für i = 0, 1, ..., 159

**[0058]** Das heißt, für jede Abtastung s'[i] in dem hochpassgefilterten Sprachsignal s' dämpft der Sprachdämpfer/Komfortgeräuscheinkoppler **270** die Abtastung s'[i] um die Dämpfung dämpf[m] des aktuellen Datenblocks.

**[0059]** Gleichzeitig koppelt der Sprachdämpfer/Komfortgeräuscheinkoppler **270** auch das tiefpassgefilterte weiße Rauschen n[i] auf der Grundlage des Wertes von dämpf[m] ein. Wie aus der obigen Gleichung zu ersehen ist, erfolgt bei dämpf[m] = 1 keine Dämpfung, und s''[i] = s'[i]. Wenn dämpf[m] = maxdämpf (0,65), dann ist s''[i] = (0,65\*hochpassgefiltertes Sprachsignal) + (0,35\*tiefpassgefiltertes weißes Rauschen). Die Dämpfung des Signals s'[i] und das Einkoppeln von tiefpassgefiltertem weißem Rauschen (Komfortgeräusch) dient dem Erzeugen eines glatteren Hintergrundgeräuschs mit weniger wahrnehmbarem Wirbel.

**[0060]** Das vom Sprachdämpfer/Komfortgeräuscheinkoppler **270** erzeugte Signal s''[i] kann dem Digital-Analog-Wandler **60** oder einer anderen Vorrichtung zugeführt werden, welche das Signal in ein anderes digitales Datenformat umwandelt, wie oben besprochen.

**[0061]** Wie oben besprochen, wirken der Dämpfungsberechner **240**, der Geräuschformungsgenerator **250** und der Sprachdämpfer/Komfortgeräuscheinkoppler **270** zum Zweck der Verringerung des Hintergrundwirbels zusammen, wenn in dem empfangenen Signal keine Sprache vorhanden ist. Diese Elemente könnten als ein einziger Geräuschminderer betrachtet werden, der in [Fig. 2](#) innerhalb der gestrichelten Linien als **280** gezeigt ist. Dieser Geräuschminderer **280** empfängt das Spracherkennungssignal v[m] vom Sprachaktivitätsdetektor **230**, die Geräuschschätzung N[m] vom Geräuschschätzer **220** und das hochpassgefilterte Signal s'[i] vom

Hochpassfiltertreiber **260** und erzeugt einen verarbeiteten dekodierten Sprachbitstrom  $s''[i]$ , wie oben besprochen.

**[0062]** Ein geeigneter VADDNR **50**, wie oben beschrieben, könnte in einem Mikroprozessor, wie in [Fig. 6](#) gezeigt, implementiert werden. Der Mikroprozessor ( $\mu$ ) **610** ist über eine Datenleitung **621** und eine Adressleitung **622** mit einem nicht-flüchtigen Speicher **620**, wie beispielsweise einem ROM, verbunden. Der nicht-flüchtige Speicher **620** enthält einen Programmcode zur Implementierung der Funktionen des VADDNR **50**, wie oben besprochen. Der Mikroprozessor **610** ist des Weiteren über eine Datenleitung **631** und eine Adressleitung **632** mit einem flüchtigen Speicher **630**, wie beispielsweise einem RAM, verbunden. Der Mikroprozessor **610** empfängt den dekodierten Sprachbitstrom  $s$  vom Sprachdekodierer **40** auf Signalleitung **612** und erzeugt einen verarbeiteten dekodierten Sprachbitstrom  $s''$ . Wie oben besprochen, wird in einer Ausführungsform der vorliegenden Erfindung der VSELP-kodierte Datenblockenergiewert  $r_0$  dem VADDNR **50** aus dem kodierten Sprachbitstrom  $b$  zugeführt. Dies ist in [Fig. 6](#) durch die Signalleitung **611** gezeigt. In einer anderen Ausführungsform berechnet der VADDNR den Datenblockenergiewert  $r_0$  aus dem dekodierten Sprachbitstrom  $s$ , und eine Signalleitung **611** wäre nicht vorhanden.

**[0063]** Es versteht sich, dass die in der vorliegenden Beschreibung gezeigten und beschriebenen Ausführungsformen und Variationen die Erfindung lediglich veranschaulichen und dass der Fachmann verschiedene Modifikationen implementieren kann.

**[0064]** In dieser Beschreibung wurden verschiedene bevorzugte Werte und Wertebereiche offenbart. Es ist jedoch zu beachten, dass diese Werte sich auf den Gebrauch der vorliegenden Erfindung in einer sich bewegenden Umgebung beziehen. Der Fachmann erkennt, dass die im vorliegenden Text offenbarte Erfindung in verschiedenen Umgebungen realisiert werden kann, wobei dann die Werte und Wertebereiche von denen, die im vorliegenden Text besprochen wurden, abweichen können. Der Geltungsbereich der Erfindung ist somit nur durch die angehängten Ansprüche begrenzt.

### Patentansprüche

1. Verfahren zum Verarbeiten eines empfangenen Signals, das Sprache und Geräusche repräsentiert, wobei das Verfahren folgende Schritte umfasst:

Erzeugen eines Energiewertes, der die akustische Energie des empfangenen Signals repräsentiert, wobei das empfangene Signal keine speziellen sprachfreien Datenblöcke enthält, wobei sprachfreie Datenblöcke als Datenblöcke, die Komfortgeräusch-Parameter enthalten, definiert sind;

Erzeugen – anhand dieses Energiewertes – eines Geräuscheschätzungswertes, der das durchschnittliche Hintergrundgeräusch in dem empfangenen Signal repräsentiert;

Erzeugen eines hochpassgefilterten Signals durch Anlegen des empfangenen Signals an eines von mehreren Hochpassfiltern anhand des Geräuscheschätzungswertes;

Erzeugen eines Komfortgeräuschs anhand des Geräuscheschätzungswertes;

Bestimmen – anhand des Energiewertes und des Geräuscheschätzungswertes –, ob das empfangene Signal eine Sprachkomponente enthält; und

Erzeugen – anhand des hochpassgefilterten Signals – eines verarbeiteten hochpassgefilterten Signals, wenn das empfangene Signal keine Sprachkomponente enthält.

2. Verfahren nach Anspruch 1, wobei die Differenz in den Grenzfrequenzen jedes der mehreren Hochpassfilter mindestens 100 Hz beträgt.

3. Verfahren nach Anspruch 1, wobei der Schritt des Erzeugens eines verarbeiteten hochpassgefilterten Signals folgende Schritte umfasst:

Dämpfen des hochpassgefilterten Signals; und

Einfügen des Komfortgeräuschs in das hochpassgefilterte Signal.

4. Verfahren nach Anspruch 3, wobei das Komfortgeräusch ein tiefpassgefiltertes weißes Rauschen umfasst, das anhand des Geräuscheschätzungswertes skaliert wird.

5. Vorrichtung zum Verarbeiten eines empfangenen Signals, wobei das empfangene Signal eine Sprachkomponente und eine Geräuschkomponente umfasst, wobei die Vorrichtung folgendes umfasst:

Mittel (**210**) zum Erzeugen eines Energiewertes ( $e(m)$ ), der die akustische Energie des empfangenen Signals repräsentiert;

Mittel (**220**) zum Erzeugen – anhand dieses Energiewertes – eines Geräuscheschätzungswertes ( $N(m)$ ), der das

durchschnittliche Hintergrundgeräusch in dem empfangenen Signal repräsentiert; und Mittel **(280)** zum Erzeugen eines hochpassgefilterten Signals durch Anlegen des empfangenen Signals an eines von mehreren Hochpassfiltern anhand des Geräuschschätzungswertes, wobei das Mittel zum Erzeugen des Energiewertes und das Mittel zum Erzeugen des Hochpassfilters sich in der Empfangsvorrichtung befinden.

6. Vorrichtung nach Anspruch 5, wobei die Differenz in den Grenzfrequenzen jedes der mehreren Hochpassfilter mindestens 100 Hz beträgt.

7. Vorrichtung nach Anspruch 5, umfassend:  
Mittel zum Bestimmen, ob das empfangene Signal eine Sprachkomponente enthält; und  
Mittel zum Erzeugen eines verarbeiteten hochpassgefilterten Signals, wenn das empfangene Signal keine Sprachkomponente enthält.

8. Vorrichtung nach Anspruch 7, wobei das Mittel zum Erzeugen eines verarbeiteten hochpassgefilterten Signals folgendes umfasst:  
Mittel **(250)** zum Erzeugen von Komfortgeräusch anhand des Geräuschschätzungswertes;  
Mittel **(270)** zum Dämpfen des hochpassgefilterten Signals; und  
Mittel **(270)** zum Einfügen des Komfortgeräuschs in das hochpassgefilterte Signal.

Es folgen 6 Blatt Zeichnungen

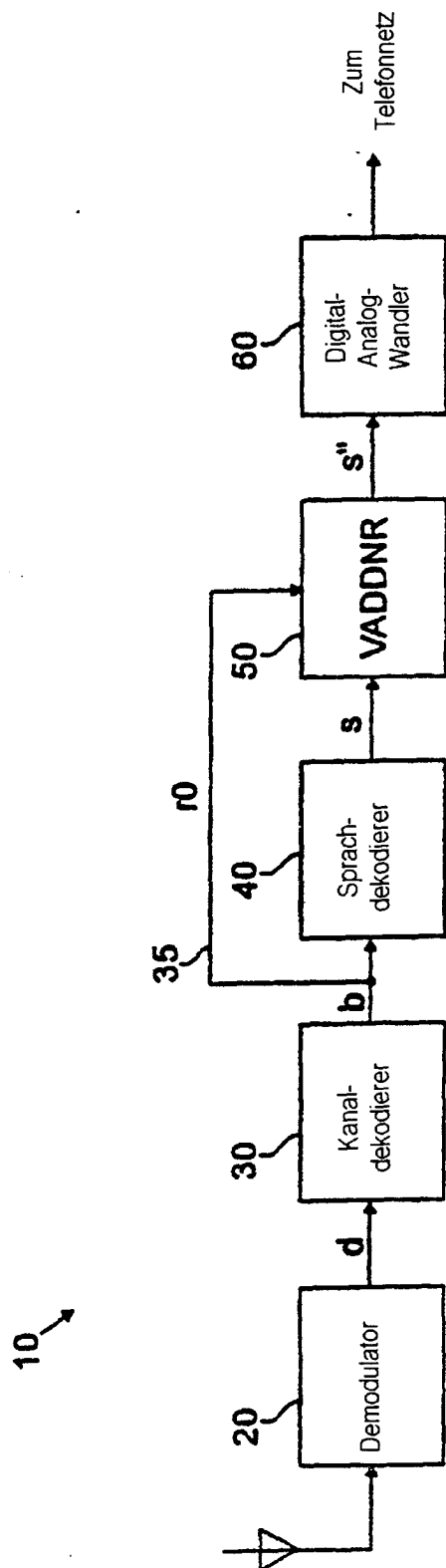


FIG. 1

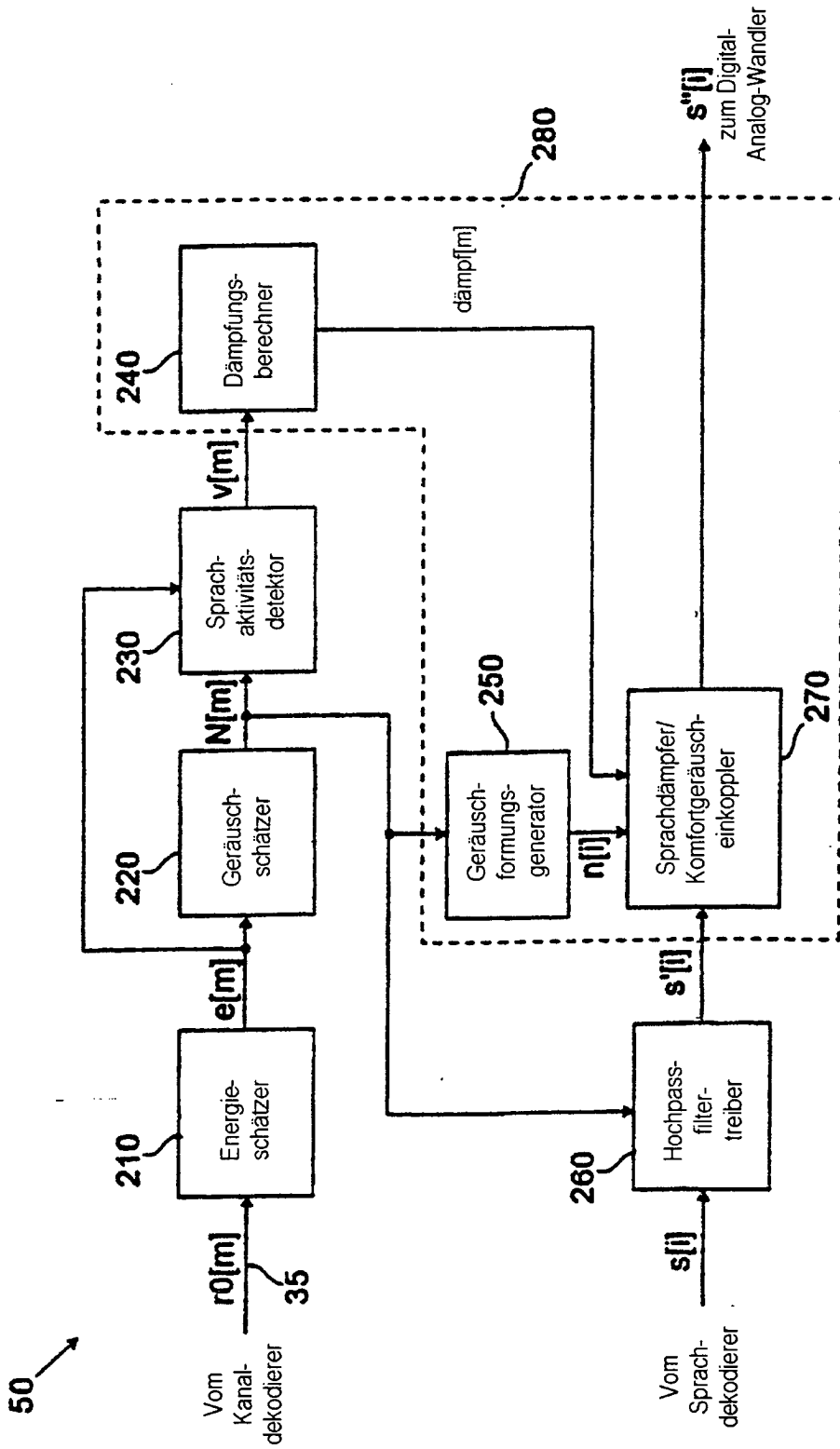


FIG. 2

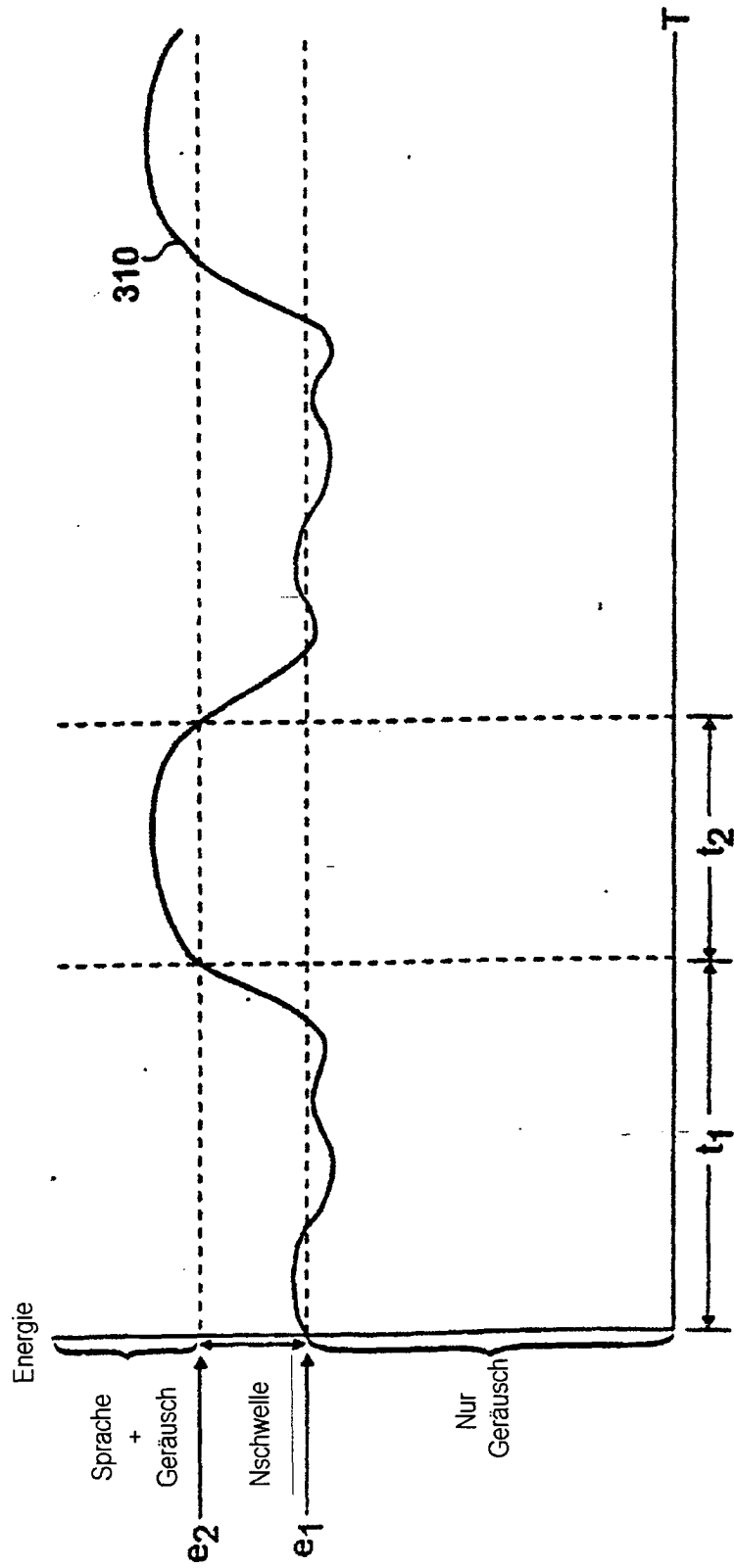


FIG. 3

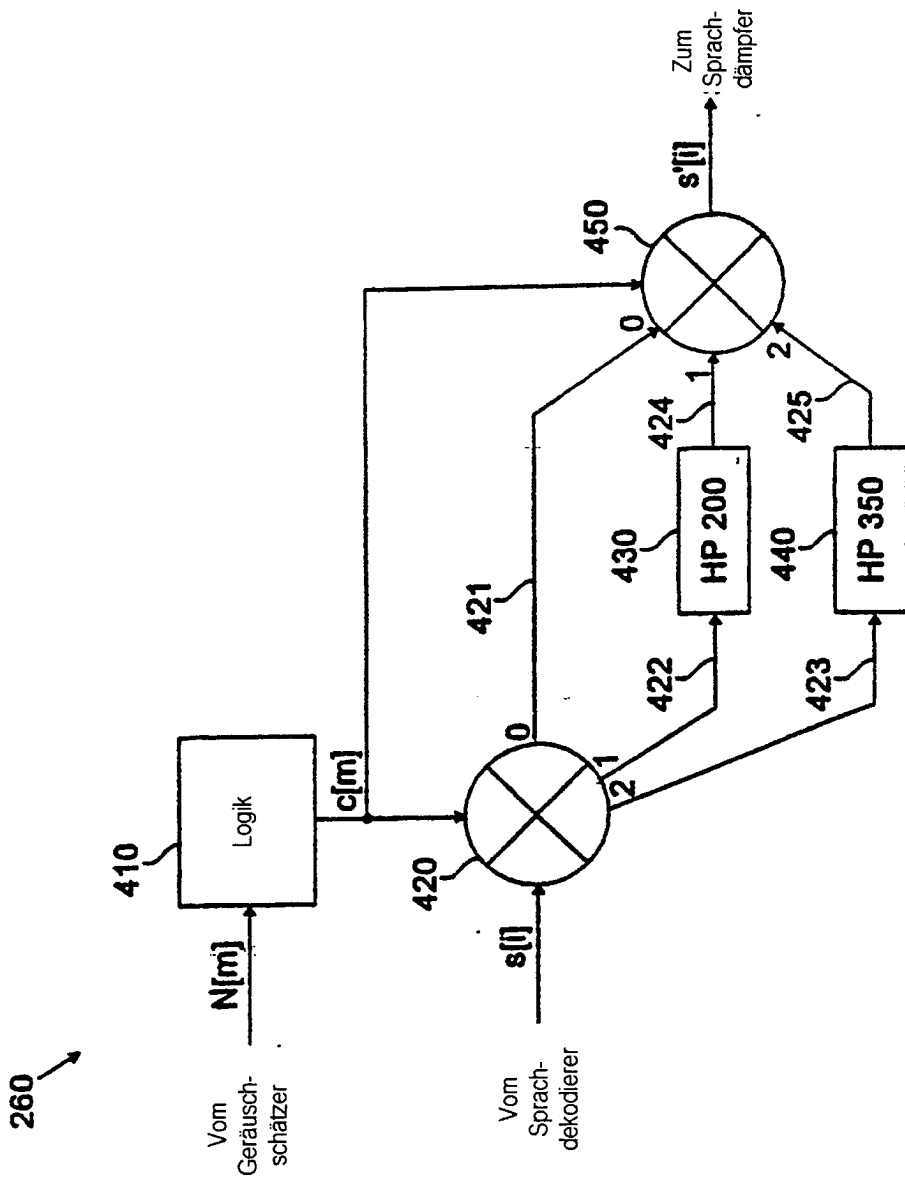


FIG. 4

260 ↗

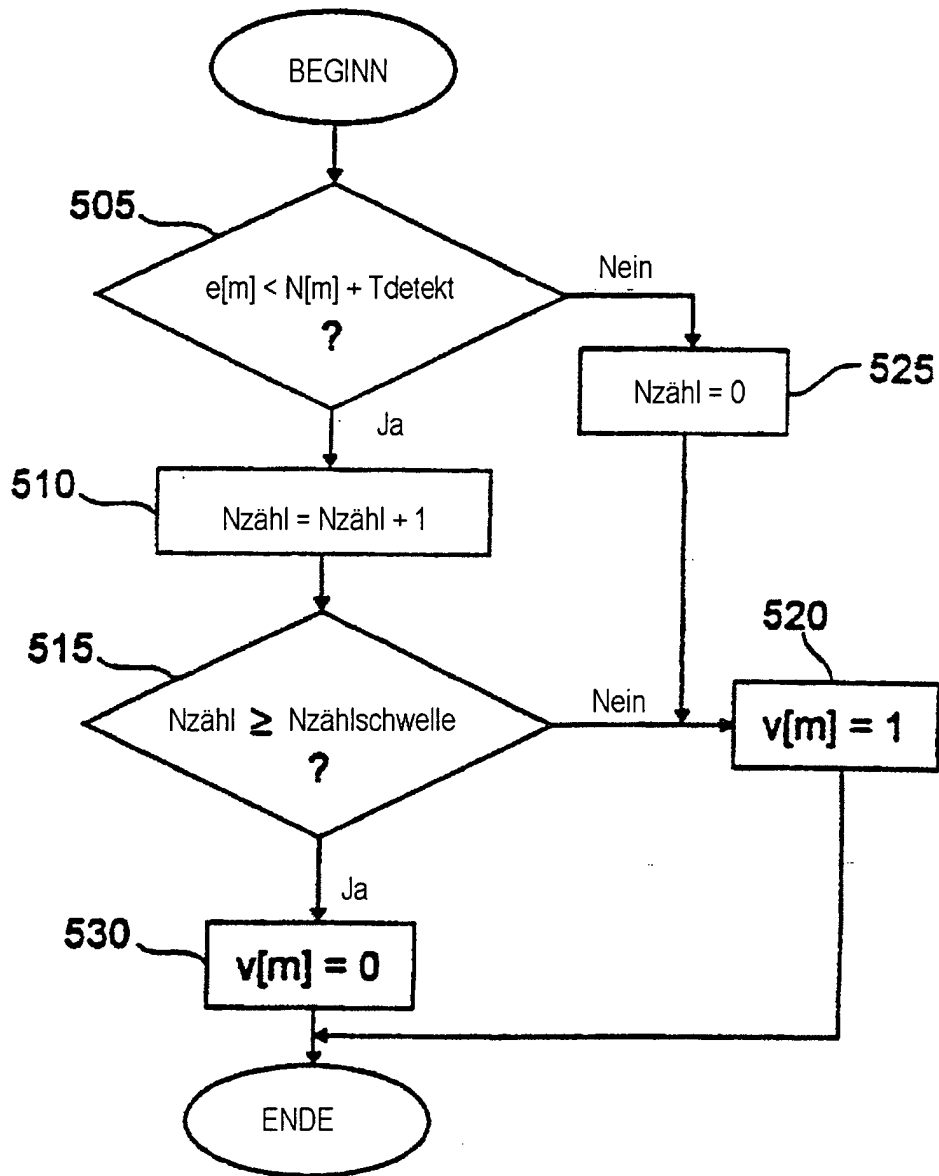
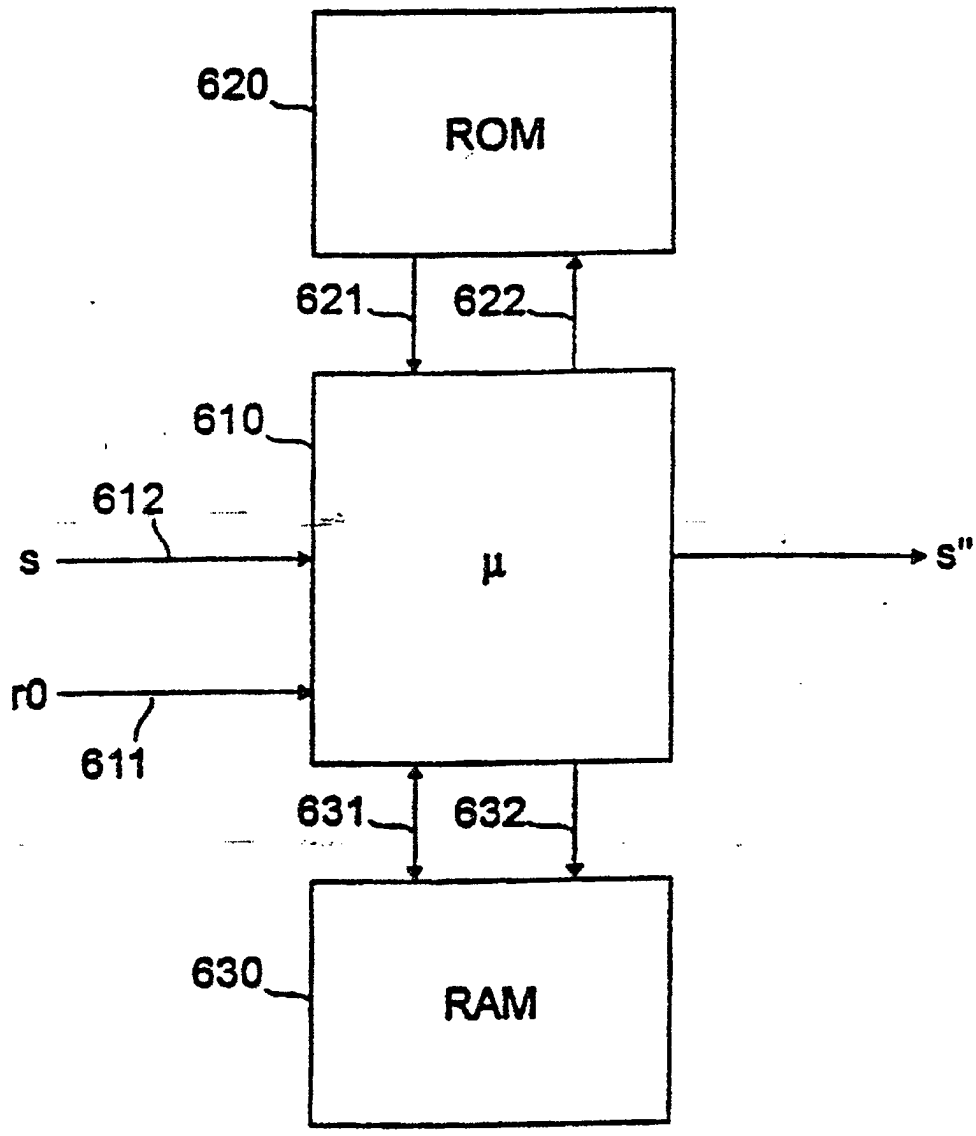


FIG. 5





**FIG. 6**