

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 874 352 A2

(12)

EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:

28.10.1998 Patentblatt 1998/44

(51) Int. Cl.⁶: **G10L 3/00**

(21) Anmeldenummer: **98102842.6**

(22) Anmeldetag: **19.02.1998**

(84) Benannte Vertragsstaaten:

**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE**

Benannte Erstreckungsstaaten:

AL LT LV MK RO SI

(30) Priorität: **22.04.1997 DE 19716862**

(71) Anmelder: **Deutsche Telekom AG**

53113 Bonn (DE)

(72) Erfinder:

• **Stegmann, Joachim, Dipl.-Ing.**
64289 Darmstadt (DE)

• **Schröder, Gerhard, Dipl.-Ing.**
64807 Dieburg (DE)

(54) Sprachaktivitätserkennung

(57) Es werden ein Verfahren und eine Schaltungsanordnung zur automatischen Sprachaktivitätserkennung auf Basis der Wavelet-Transformation angegeben. Zur quellengesteuerten Reduktion der mittleren Übertragungsrate wird eine Sprachaktivitätserkennungsschaltung bzw. ein -modul zur Steuerung eines Sprachcodierers und eines Sprachdecodierers und zur Steuerung eines Hintergrundgeräuschcodierers sowie eines Hintergrundgeräuschdecodierers verwendet. Nach der Segmentierung eines Sprachsignals wird für jeden Rahmen eine Wavelet-Transformation berechnet, aus der ein Satz Parameter ermittelt wird, aus denen wiederum mit Hilfe fester Schwellen ein Satz binärer Entscheidungsvariablen in einer Rechenschaltung (32) berechnet wird. Die Entscheidungsvariablen steuern eine Entscheidungslogik (42), deren Resultat nach zeitlicher Glättung in einer Schaltung (44) für jeden Rahmen eine Aussage "Sprache vorhanden / keine Sprache" liefert. Die Schaltung selbst besteht im wesentlichen aus einer Segmentierungsschaltung (28), einer Wavelet-Transformationsschaltung (30), einer Rechenschaltung für die Energiegrößen (32), einer Schaltung für Pausendetektion (34), einer Schaltung für das Stationaritätsmaß (35), einem ersten und einem zweiten Hintergrunddetektor (36 bzw. 37), einer nachgeschalteten Entscheidungslogik (42) und der Schaltung (44) für die zeitliche Glättung, die an ihrem Ausgang (45) die gewünschte Aussage liefert.

EP 0 874 352 A2

Beschreibung

Die Erfindung betrifft ein Verfahren und eine Schaltungsanordnung zur automatischen Sprachaktivitätserkennung nach dem Oberbegriff des Patentanspruchs 1 bzw. 5.

Für den digitalen Mobilfunk oder Sprachspeichersysteme und eine große Anzahl anderer Anwendungen ist es vorteilhaft, eine diskontinuierliche Übertragung der Sprachcodierungsparameter vorzunehmen. Dadurch kann während der Sprachpausen oder Zeitintervallen, die im wesentlichen durch Hintergrundgeräusche dominiert werden, die Bitrate deutlich gesenkt werden. Vorteile ergeben sich daraus unter anderem durch einen niedrigen Energieverbrauch in mobilen Endgeräten, durch eine höhere mittlere Bitrate für simultane Dienste, wie Datenübertragung oder durch eine höhere Speicherkapazität auf Speicherchips. Das Ausmaß der Vorteile hängt vom Anteil der Pausen im Sprachsignal sowie von der Güte des automatischen Sprachaktivitätserkenners ab, der zur Detektion solcher Intervalle nötig ist. Zum einen ist eine niedrige Sprachaktivitätsrate vorteilhaft, zum anderen sollte aktive Sprache nicht abgeschnitten werden, um die Sprachqualität nicht zu beeinträchtigen. Dieses Problem ist das grundsätzliche Problem von automatischen Sprachaktivitätserkennern, besonders bei der Präsenz von Hintergrundgeräuschen hohen Pegels.

Bekannte Verfahren zur automatischen Sprachaktivitätserkennung benutzen in der Regel Entscheidungsparameter, die auf Zeitmittelwerten über Fenster konstanter Länge basieren. Als Beispiel hierfür werden die Autokorrelationskoeffizienten, Nulldurchgangsraten oder Sprachgrundperiode genannt, wobei diese Parameter nur eine beschränkte Flexibilität in der Auswahl der Zeit-/Frequenzbereichsauflösung besitzen, welche normalerweise durch die Rahmenlänge des zugehörigen Sprachcodierers/-decodierers fest vorgegeben ist. Im Gegensatz dazu berechnet die bekannte Wavelet-Transformation eine Zerlegung in den Zeit-/Frequenzbereich der zu niedriger Frequenz- aber hoher Zeitbereichsauflösung bei hohen Frequenzen und zu niedriger Zeit- aber hoher Frequenzbereichsauflösung bei niedrigen Frequenzen führt. Für die Analyse von Sprachsignalen sind diese Eigenschaften gut geeignet und deshalb zur Klassifizierung von aktiver Sprache in die Klassen stimmhaft, stimmlos und Übergänge bereits genutzt worden, wie in der Offenlegungsschrift DE 195 38 852 A1 "Verfahren und Anordnung zur Klassifizierung von Sprachsignalen", 1997, ausgeführt wird.

Der Erfindung liegt deshalb die Aufgabe zugrunde, ein Verfahren und eine Schaltungsanordnung zur Sprachaktivitätserkennung auf Basis der Wavelet-Transformation zu schaffen, wobei entschieden werden soll, ob für das zu betrachtende Zeitsegment überhaupt Sprache oder Sprachlaute vorliegen.

Die erfindungsgemäße Lösung des Verfahrens ist im Kennzeichen des Patentanspruchs 1 charakterisiert.

Weitere Ausgestaltungen des erfindungsgemäßen Verfahrens ergeben sich aus den Patentansprüchen 2 bis 4.

Die Lösung für die erfindungsgemäße Schaltungsanordnung ist in den Kennzeichen des Patentanspruchs 5 charakterisiert.

Weitere Merkmale bzw. Ausgestaltungen der Schaltungsanordnungen sind im Patentanspruch 6 charakterisiert.

Das vorliegende Verfahren zur automatischen Sprachaktivitätserkennung für Sprachcodierer/-decodierer zur quellengesteuerten Reduktion der mittleren Übertragungsrate zeichnet sich dadurch aus, daß nach der Segmentierung des Sprachsignals für jeden Rahmen eine Wavelet-Transformation berechnet wird, aus der ein Satz Parameter ermittelt wird, aus denen mit Hilfe fester Schwellen ein Satz binärer Entscheidungsvariablen berechnet wird, die eine Entscheidungslogik steuern, deren Resultat nach zeitlicher Glättung für jeden Rahmen eine Aussage "Sprache vorhanden / keine Sprache vorhanden" liefert. Dadurch, daß hiermit festgestellt wird, ob für das zu betrachtende Zeitsegment überhaupt Sprache vorliegt, wird eine quellengesteuerte Reduktion der mittleren Übertragungsrate erreicht.

Weitere Vorteile des Verfahrens zur Sprachaktivitätserkennung sowie der entsprechenden Schaltungsanordnung werden anhand von Ausführungsbeispielen, die im folgenden näher beschrieben werden, aufgezeigt.

In der Beschreibung, in den Patentansprüchen, in der Zusammenfassung und in der Zeichnung werden die in der hinten angeführten Liste der Bezugszeichen verwendeten Begriffe und Bezugszeichen verwendet.

Die Erfindung wird nun anhand von Ausführungsbeispielen, die in den Zeichnungen dargestellt sind, beschrieben. In der Zeichnung bedeuten:

Fig. 1 ein Blockschaltbild für die Sprachaktivitätserkennung als Vorstufe für einen variabelratigen Sprachcodierer/-decodierer und

Fig. 2 ein Blockschaltbild eines automatischen Sprachaktivitätserkenners.

Durch das Verfahren wird entschieden, ob für das zu betrachtende Zeitsegment überhaupt Sprache vorliegt. Dadurch kann es zur Funktionssteuerung oder allgemein als Vorstufe für einen Sprachcodierer/-decodierer mit variabler Bitrate verwendet werden. In Fig. 1 ist ein Blockschaltbild für die Sprachaktivitätserkennung als Vorstufe für einen variabelratigen Sprachcodierer/-decodierer gezeigt, der auf seinem Eingang 1 die jeweilige Eingangssprache empfängt. Über Leitungen 2 und 3 wird die Eingangssprache sowohl auf einen Umschalter 4 als auch auf den Eingang einer Sprachaktivitätserkennungsschaltung bzw. ein Modul 5 gegeben. Der Umschalter 4 leitet die Eingangssprache abhängig vom Ausgangssignal der Sprachaktivitätserkennungsschaltung 5, die zur Steuerung des Umschalters 4 über eine

Rückmeldeleitung 6 mit ihm verbunden ist, entweder auf die Leitung 7 oder auf die Leitung 8. Die Leitung 7 führt auf einen Sprachcodierer 9 und die Leitung 8 auf einen Hintergrundgeräuschcodierer 10. Der Bitstrom des Sprachcodierers 9 wird über eine Leitung 11 auf den einen Eingang eines Umschalters 13 gegeben und der Bitstrom des Hintergrundgeräuschcodierers 10 über eine Leitung 12 auf den anderen Eingang des Umschalters 13. Der Umschalter 13 empfängt über eine Leitung 14 auch die Ausgangssignale der Sprachaktivitätserkennungsschaltung 5, wodurch der Umschalter 13 gesteuert wird. Der Ausgang des Umschalters 13 ist über eine Leitung 15 mit einem Übertragungskanal 16 verbunden, der eingangsseitig auch noch mit der Leitung 14 für die Ausgangssignale des Sprachaktivitätserkennungsmoduls 5 verbunden ist. Der Ausgang des Übertragungskanals 16 gelangt einmal über eine Leitung 17 auf den Eingang eines weiteren Umschalters 19 und über eine Leitung 18 auf den Steuereingang des Umschalters 19 und den Steuereingang eines Umschalters 26. Der Umschalter 19 ist über Ausgangsleitungen 20 und 21 mit einem Sprachdekodierer 22 und mit einem Dekodierer für Hintergrundgeräusche 23 verbunden, deren Ausgänge über Leitungen 24 bzw. 25 auf den Eingang des bereits genannten Umschalters 26 gelangen, der in Abhängigkeit der Steuersignale auf Leitung 18 am Ausgang 27 entweder Signale für die decodierte Sprache oder die decodierten Hintergrundgeräusche bereitstellt.

In Fig. 2 ist ein Blockschaltbild eines automatischen Sprachaktivitätserkenners dargestellt, der wiederum auf seinem Eingang 1 die Eingangssprache empfängt und an eine Segmentierungsschaltung 28 weiterleitet. Der Ausgang dieser Segmentierungsschaltung 28 wird über eine Leitung 29 an eine Wavelet-Transformationsschaltung 30 übertragen, die wiederum über eine Leitung 31 mit dem Eingang eines Prozessors 32 zur Berechnung der Energiegrößen verbunden ist. Der Ausgang des Prozessors 32 bzw. einer Rechenschaltung ist über eine Leitung 33 parallel mit einem Pausendetektor 34, mit einer Schaltung zur Berechnung eines Stationaritätsmaßes 35, mit einem ersten Hintergrunddetektor 36 und mit einem zweiten Hintergrunddetektor 37 verbunden. Die Ausgänge der genannten Schaltungen 34 bis 37 sind über entsprechende Leitungen 38 bis 41 mit einer Entscheidungslogik 42 verbunden, deren Ausgang über eine Leitung 43 mit einer Glättungsschaltung 44 zur zeitlichen Glättung verbunden ist, deren Ausgang 45 auch Ausgang des Sprachaktivitätserkenners ist.

Anhand des Blockschaltbildes des automatischen Sprachaktivitätserkenners nach Fig. 2 wird nun das Verfahren zur automatischen Sprachaktivitätserkennung näher beschrieben. Nach der Segmentierung des Eingangssignals in der Segmentierungsschaltung 28 wird für jedes Segment die Wavelet-Transformation in der Wavelet-Transformationschaltung 30 berechnet. Im Anschluß daran wird ein Satz von Energieparametern im Prozessor 32 aus den Transformationskoeffizienten ermittelt und mit festen Schwellwerten verglichen. Dadurch entstehen binäre Entscheidungsvariablen, mit denen die Entscheidungslogik 42 gesteuert wird, die ein vorläufiges Resultat für jeden Rahmen ausgibt. Diese vorläufige Entscheidung wird zum Abschluß noch mittels zeitlicher Glättung in der Schaltung 44 nachverarbeitet, wodurch das endgültige Resultat "Sprache oder keine Sprache" am Ausgang 45 für den aktuellen Rahmen zur Verfügung steht. Im nachfolgenden werden nun die einzelnen Verarbeitungsschaltungen bzw. -blöcke des Blockdiagramms nach Fig. 2 näher beschrieben. Die Wavelet-Transformationsschaltung 30 führt folgende Operationen aus: Die Eingangssprache wird in Rahmen der Länge N Abtastwerte eingeteilt, wobei N an ein bestimmtes Sprachcodierverfahren angepaßt werden kann. Für jeden Rahmen wird die diskrete Wavelet-Transformation berechnet. Für viele Basisfunktionen der Wavelet-Transformation kann eine Filterbank mit jeweils einem Hochpaßfilter oder einem Tiefpaßfilter abgeleitet werden, mit denen die Transformation rekursiv durchgeführt werden kann. Auf die Klassen der Daubechies Wavelets und der Spline Wavelets wird hingewiesen, die zu einer besonders effizienten Implementierung der Transformation mittels Filter kurzer Länge führen. Im nachfolgenden werden zwei Verfahren beschrieben, die sich zur Berechnung der Transformation eignen.

1. Im ersten Verfahren wird die Filterbank direkt auf den Eingangs-Sprachrahmen $\underline{s} = (s(0), \dots, s(N-1))^T$ angewandt und die beiden Filterausgänge um den Faktor zwei unterabgetastet. Dadurch entsteht am Tiefpaßfilterausgang ein Satz von Approximationskoeffizienten $\underline{A}_1 = (A_1(0), \dots, A_1(N/2-1))^T$ sowie am Hochpaßfilterausgang ein Satz von Detailkoeffizienten $\underline{D}_1 = (D_1(0), \dots, D_1(N/2-1))^T$. Dieses Verfahren wird nun rekursiv immer auf die Approximationskoeffizienten der vorherigen Stufe angewandt, wodurch sich in der letzten Stufe L als Ergebnis der Transformation der Vektor $\text{DWT}(\underline{s}) = (\underline{D}_1^T, \underline{D}_2^T, \dots, \underline{D}_L^T, \underline{A}_L^T)^T$ mit insgesamt N Koeffizienten ergibt.

2. Die zweite Variante zur Berechnung der Transformation beruht wie die erste auf einer Filterbankzerlegung. Der Unterschied besteht aber darin, daß die Filterausgänge nicht mehr unterabgetastet werden. Dadurch entstehen nach jeder Stufe Vektoren der Länge N und nach der letzten Stufe ein Ausgangsvektor mit insgesamt $(L+1)N$ Koeffizienten. Um die Auflösungseigenschaften der Wavelet-Transformation zu erhalten, müssen die Filterimpulsantworten für jede Stufe durch Überabtastung um den Faktor zwei aus der vorherigen Stufe gewonnen werden. In der ersten Stufe werden die gleichen Filter benutzt wie für Variante 1. Durch den insgesamt höheren Aufwand - höhere Redundanz in der Darstellung im Bildbereich - gegenüber Variante 1 kann eine Verbesserung der Leistungsfähigkeit des Verfahrens erreicht werden.

Um Randeffekte aufgrund der Filterlänge M zu vermeiden, werden die $M \cdot 2^{L-2}$ vergangenen sowie die $M \cdot 2^{L-2}$ zukünftigen Abtastwerte des Sprachrahmens ebenfalls berücksichtigt und die Filter-Impulsantworten - soweit wie möglich - um den zeitlichen Ursprung zentriert. Dadurch erhöht sich die algorithmische Verzögerung des Verfahrens prinzipiell um $M \cdot 2^{L-2}$ Abtastwerte. Soll dies vermieden werden, so kann der Eingangsrahmen alternativ auch periodisch oder symmetrisch fortgesetzt werden.

Zunächst werden die Rahmenenergien $E_1, \dots, E_L$ der Detailkoeffizienten $\underline{D}_1, \dots, \underline{D}_L$ und die Rahmenenergie E_{L+1} der Approximationskoeffizienten $\underline{A}_L$ vom Prozessor 32 berechnet.

Die Gesamtenergie des Rahmens E_{tot} kann man nun effizient durch Summation aller partieller Energien ermitteln, wenn die zugrundeliegende Wavelet-Basis orthogonal ist. Alle Energiewerte werden im logarithmischen Bereich repräsentiert.

Für die Pausendetektion in der Schaltung 34 wird die Rahmenenergie E_{tot} mit einer festen Schwelle T_1 verglichen, um Rahmen mit sehr niedriger Energie zu erkennen. Dazu wird eine binäre Entscheidungsvariable f_{sil} gemäß der folgenden Formel gesetzt:

$$f_{\text{sil}} = \begin{cases} 1, & E_{\text{tot}} < T_1 \\ 0, & \text{sonst} \end{cases} \quad (1)$$

Um bei der Detektion von stationären Rahmen ein Maß für stationäre oder instationäre Rahmen zu erhalten, wird für jeden Rahmen k das Differenzmaß

$$\Delta^{(k)} = \sqrt{\frac{1}{L} \sum_{i=1}^L (E_i^{(k)} - E_i^{(k-1)})^2} \quad (2)$$

berechnet, in das die Rahmenenergien der Detailkoeffizienten aller Stufen einfließen. Die binäre Entscheidungsvariable f_{stat} wird nun unter Verwendung der Schwelle T_2 unter Berücksichtigung der letzten K Rahmen gesetzt:

$$f_{\text{stat}} = \begin{cases} 1, & (\Delta^{(k)} < T_2 \& \dots \& (\Delta^{(k-K)} < T_2)) \\ 0, & \text{sonst} \end{cases} \quad (3)$$

Bei der Detektion von Hintergrundrauschen in den Schaltungen 36 und 37 ist das Ziel, ein Entscheidungskriterium zu erhalten, das unempfindlich gegenüber dem momentanen Pegel des Hintergrundgeräusches ist. Die Eigenschaften der DWT oder Wavelet-Transformationsschaltung 30 werden hierzu effizient ausgenutzt, indem die Detailkoeffizienten $\underline{D}_{Q1}$ im groben Zeitintervall N und die Detailkoeffizienten $\underline{D}_{Q2}$ im feineren Zeitintervall N/P betrachtet werden. Dabei bezeichnen P die Anzahl der Subrahmen, $Q1$ eine Stufe für grobe und $Q2$ eine Stufe für feine Zeitauflösung, wobei die Beziehungen $Q1, Q2 \in \{1, L\}$ sowie $Q1 > Q2$ gelten müssen. Vorab wird für beide Stufen ein Schätzwert $B_i, i \in \{Q1, Q2\}$, für den momentanen Pegel des Hintergrundgeräusches durch

$$B_i^{(k)} = \begin{cases} E_i^{(k)}, & B_i^{(k-1)} > E_i^{(k)} \\ \alpha B_i^{(k-1)} + (1-\alpha) E_i^{(k)}, & \text{sonst} \end{cases} \quad (4)$$

berechnet, wobei die Zeitkonstante α durch $0 < \alpha < 1$ begrenzt ist. Danach werden die P Subrahmenenergien

$$\in \begin{matrix} (k,1) \\ Q2 \end{matrix}, \dots, \in \begin{matrix} (k,P) \\ Q2 \end{matrix}$$

aus den Detailkoeffizienten $\underline{D}_2$ bestimmt und mit Hilfe der festen Schwellen T_3 und T_4 jeweils eine binäre Entscheidungsvariable f_{Q1} für die Stufe Q1 sowie f_{Q2} für die Stufe Q2 gemäß den folgenden beiden Formeln bestimmt:

(5)

$$f_{Q1} = \begin{cases} 1, & (E_{Q1}^{(k)} - B_{Q1}^{(k)}) < T_3 \\ 0, & \text{sonst} \end{cases}$$

$$f_{Q2} = \begin{cases} 1, & [(\in_{Q2}^{(k,1)} - B_{Q2}^{(k)}) < T_4] \dots [(\in_{Q2}^{(k,P)} - B_{Q2}^{(k)}) < T_4] \\ \text{sonst} \end{cases}$$

(6)

In der Entscheidungslogik 42 wird unter Verwendung der Gleichungen (1), (3), (5) und (6) das vorläufige Resultat $\text{vad}^{(\text{pre})}$ des automatischen Sprachaktivitätserkenners durch die logische Verknüpfung

$$\text{vad}^{(\text{pre})} = !(\text{f}_{\text{sil}} | \text{f}_{Q1} \& \text{f}_{Q2} \& \text{f}_{\text{stat}}), \quad (7)$$

gewonnen, wobei '!', '&' die logischen Operatoren 'nicht', 'oder', und 'und' bezeichnen.

Es können auch weitere Stufen Q3, Q4, ..., etc. definiert werden, für die eine Detektion des Hintergrundrauschens in gleicher Weise vorgenommen werden kann. Es sind dann weitere binäre Entscheidungsparameter f_{Q3} , f_{Q4} , ... zu definieren, die ebenso in Gleichung (7) zu berücksichtigen sind.

Die zeitliche Glättung erfolgt in der Schaltung 44. Um die Langzeit-Stationarität von Sprache in Betracht zu ziehen, wird die vorläufige Entscheidung der VAD in einer nachverarbeitenden Stufe noch zeitlich geglättet. Übersteigt die Anzahl der letzten zusammenhängend als aktiv gekennzeichneten Rahmen den Wert C_B , so werden, solange $\text{vad}^{(\text{pre})} = 0$, noch maximal C_H aktive Rahmen angehängt. Damit liegt die endgültige Entscheidung $\text{vad} \in \{0,1\}$ des Sprachaktivitätserkenners fest.

Liste der Bezugszeichen

1	Eingang
2,3	Leitungen
4	Umschalter
5	Sprachaktivitätserkennungsmodul oder -schaltung
6	Rückmeldeleitung
7,8	Leitungen bzw. Ausgänge des Umschalters 4
9	Sprachcodierer
10	Hintergrundgeräuschcodierer
11,12	Leitungen
13	Umschalter
14,15	Leitungen
16	Übertragungskanal
17,18	Leitungen
19	Umschalter

20,21	Leitungen
22	Sprachdecodierer
23	Hintergrundgeräuschdecodierer
24,25	Leitungen
5 26	Umschalter
27	Ausgang
28	Segmentierer
29,31,33	Leitungen
30	Wavelet-Transformationsschaltung 32 Prozessor
10 34	Phasendetektor
35	Schaltung zur Ermittlung für das Stationaritätsmaß
36,37	Hintergrunddetektor
38 - 41	Leitungen
42	Entscheidungslogik
15 43	Leitung
44	Glättungsschaltung
45	Ausgang

Patentansprüche

- 20 1. Verfahren zur automatischen Sprachaktivitätserkennung auf Basis der Wavelet-Transformation, dadurch gekennzeichnet,
- 25 daß zur quellengesteuerten Reduktion der mittleren Übertragungsrate eine Sprachaktivitätserkennungsschaltung bzw. ein -modul (5) zur Steuerung eines Sprachcodierers (7) und eines Sprachdecodierers (22) sowie zur Steuerung eines Hintergrundgeräuschcodierers (10) und eines Hintergrundgeräuschdecodierers (23) dient, wobei nach der Segmentierung eines Sprachsignals für jeden Rahmen eine Wavelet-Transformation berechnet wird, aus der ein Satz Parameter ermittelt wird, aus denen mit Hilfe fester Schwellen ein Satz binärer Entscheidungsvariablen in einer Rechenschaltung oder einem Prozessor (32) berechnet wird, die eine Entscheidungslogik (42) steuern, deren Resultat nach zeitlicher Glättung für jeden Rahmen eine Aussage "Sprache vorhanden / keine Sprache" liefert.
- 30 2. Verfahren zur Sprachaktivitätserkennung nach Patentanspruch 1, dadurch gekennzeichnet,
- 35 daß nach der Wavelet-Transformation für jedes Segment ein Satz von Energieparametern aus den Transformationskoeffizienten ermittelt und mit festen Schwellwerten verglichen wird, wodurch binäre Entscheidungsvariablen entstehen, mit denen die Entscheidungslogik (42) gesteuert wird, die ein vorläufiges Resultat für jeden Rahmen am Ausgang abgibt.
- 40 3. Verfahren zur Sprachaktivitätserkennung nach einem der Patentansprüche 1 oder 2, dadurch gekennzeichnet,
- daß das vorläufige Resultat für jeden Rahmen, das von der Entscheidungslogik ermittelt wird, mittels zeitlicher Glättung nachverarbeitet wird, wodurch das endgültige Resultat "Sprache vorhanden oder keine Sprache" für den jeweils aktuellen Rahmen gebildet wird.
- 45 4. Verfahren zur Sprachaktivitätserkennung nach einem der Patentansprüche 1 bis 3, dadurch gekennzeichnet,
- daß zur Detektion von Hintergrundrauschen Hintergrundgeräuschdetektoren (36 und 37) mit Signalen gesteuert werden und die Detailkoeffizienten (D) im groben Zeitintervall (N) und Detailkoeffizienten (D2) im feineren Zeitintervall (N/P) analysiert werden, wobei P die Anzahl der Subrahmen darstellt und die Beziehungen $Q1, Q2 \in \{1, L\}$ sowie $Q1 > Q2$ gelten.
- 50 5. Schaltungsanordnung zur Durchführung des Verfahrens zur Sprachaktivitätserkennung nach einem der Patentansprüche 1 bis 4, dadurch gekennzeichnet,
- 55 daß auf den Eingang (1) eines Umschalters (4) die Signale der Eingangssprache gelangen,
- daß mit dem Eingang (1) eine Sprachaktivitätserkennungsschaltung bzw. ein -modul (5) verbunden ist, deren

Ausgang den genannten Umschalter (4), einen weiteren Umschalter (13) steuert und außerdem mit einem Übertragungskanal (16) verbunden ist,

daß der Ausgang des Umschalters (4) über Leitungen (7 bzw. 8) mit einem Sprachcodierer (9) bzw. mit einem Hintergrundgeräuschcodierer (10) verbunden ist, deren Ausgänge über Leitungen (11 bzw. 12) mit den Eingängen des Umschalters (13) verbunden sind, dessen Ausgang über eine Leitung (15) mit dem Eingang des Übertragungskanals (16) verbunden ist, der einerseits mit einem weiteren Umschalter (19) und andererseits über eine Leitung (18) zur Steuerung des Umschalters (19) und zur Steuerung eines am Ausgang (27) angeordneten Umschalters (26) verbunden ist, und

daß zwischen den beiden Umschaltern (19 bzw. 26) ein Sprachdecodierer (22) und ein Decodierer (23) für Hintergrundgeräusche angeordnet ist.

6. Schaltungsanordnung zur Durchführung des Verfahrens nach einem der Patentansprüche 1 bis 4, dadurch gekennzeichnet,

daß der Eingang (1) mit einer Segmentierungsschaltung (28) verbunden ist, deren Ausgang über eine Leitung (29) mit einer Wavelet-Transformationsschaltung (30) verbunden ist, die mit dem Eingang einer Rechenschaltung bzw. eines Prozessors (32) zur Berechnung der Energiegrößen verbunden ist,

daß der Ausgang des Prozessors (32) über eine Leitung (33) parallel mit einem Pausendetektor (34), mit einer Schaltung zur Berechnung eines Stationaritätsmaßes (35), mit einem ersten Hintergrunddetektor (36) und mit einem zweiten Hintergrunddetektor (37) verbunden ist,

daß die Ausgänge der genannten Schaltungen (34 bis 37) mit einer Entscheidungslogik (42) verbunden sind, deren Ausgang mit einer Glättungsschaltung (44) zur zeitlichen Glättung verbunden ist, und

daß der Ausgang der Glättungsschaltung (44) auch der Ausgang (45) des Sprachaktivitätserkenners ist.

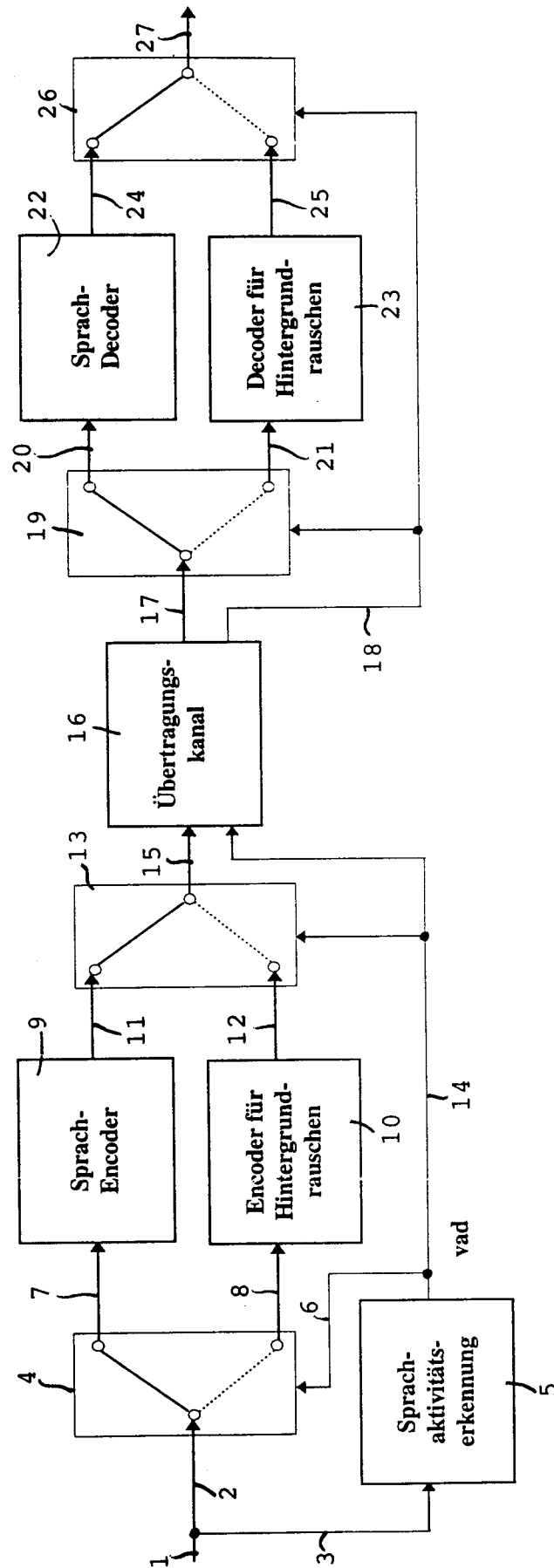


FIG. 1

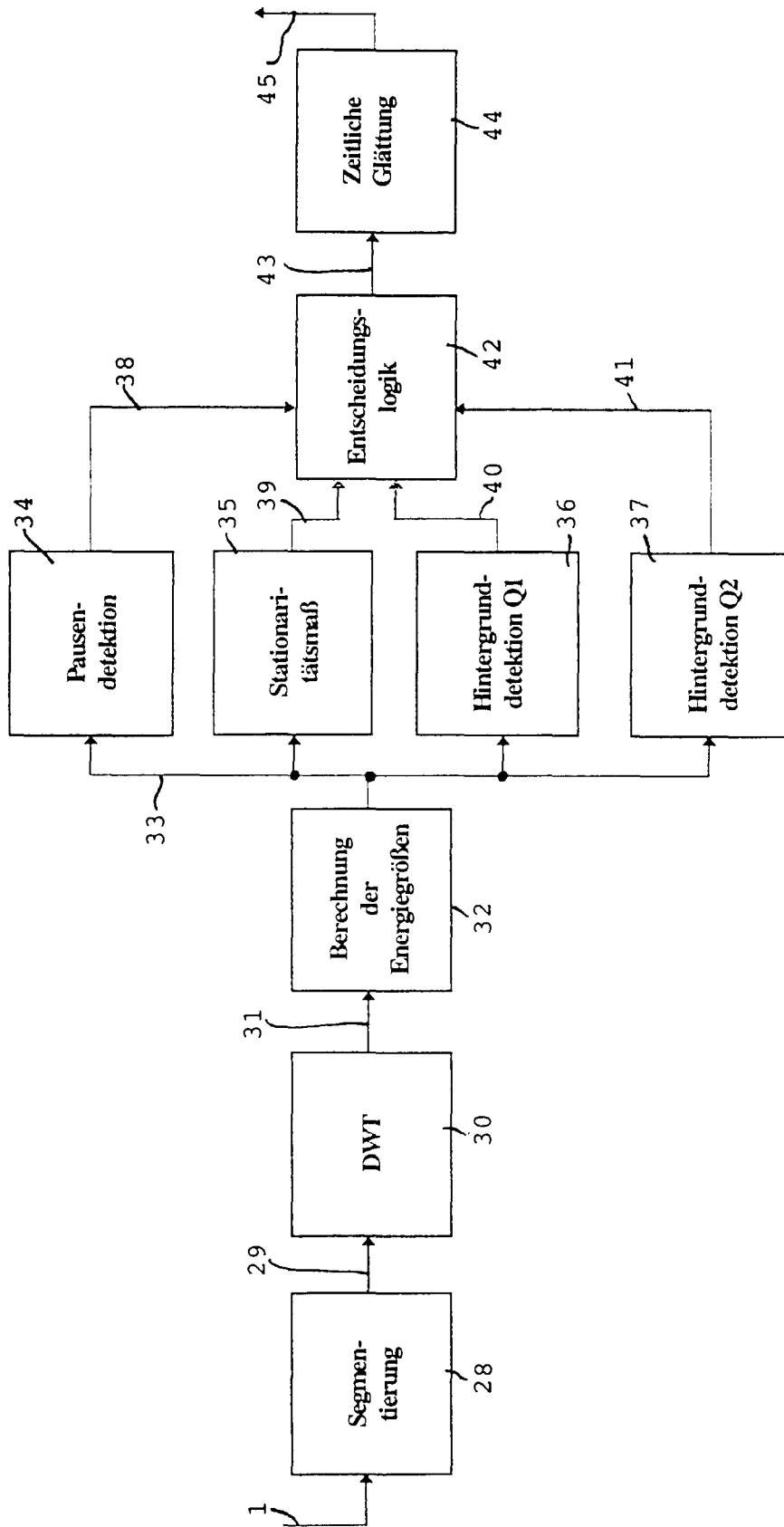


FIG. 2