

12 **EUROPÄISCHE PATENTANMELDUNG**

21 Anmeldenummer: **88202629.7**

51 Int. Cl.⁴: **G10L 3/00**

22 Anmeldetag: **23.11.88**

30 Priorität: **24.11.87 DE 3739681**

43 Veröffentlichungstag der Anmeldung:
07.06.89 Patentblatt 89/23

64 Benannte Vertragsstaaten:
DE FR GB

71 Anmelder: **Philips Patentverwaltung GmbH**
Wendenstrasse 35 Postfach 10 51 49
D-2000 Hamburg 1(DE)

64 **DE**

Anmelder: **N.V. Philips' Gloeilampenfabrieken**
Groenewoudseweg 1
NL-5621 BA Eindhoven(NL)

64 **FR GB**

72 Erfinder: **Mergel, Dieter, Dr.**
Nordalbingerweg 50a
D-2000 Hamburg 61(DE)
Erfinder: **Ney, Hermann, Dr.**
Halstenbeker Weg 65
D-2000 Hamburg 54(DE)
Erfinder: **Tomaschewski, Horst**
Am Kamp 9
D-2359 Stuenborn(DE)

74 Vertreter: **Poddig, Dieter, Dipl.-Ing. et al**
Philips Patentverwaltung GmbH
Wendenstrasse 35 Postfach 10 51 49
D-2000 Hamburg 1(DE)

54 **Verfahren zum Bestimmen von Anfangs- und Endpunkt isoliert gesprochener Wörter in einem Sprachsignal und Anordnung zur Durchführung des Verfahrens.**

EP 0 319 078 A2 57 Zur Ermittlung von Anfangs- und Endpunkt eines Wortsignals innerhalb eines Sprachsignals aus isoliert gesprochenen Wörtern werden bei jedem neuen Digitalwert drei benachbarte Fenster für die letzten bisher eingetroffenen gespeicherten Digitalwerte bestimmt, von denen das mittlere Fenster das eigentliche Wortsignal enthalten soll. Die Länge dieses mittleren Fensters wird für jeden Digitalwert zwischen einem minimalen und einem maximalen Wert variiert, und von der darin enthaltenen Energie wird jeweils ein Schwellwert subtrahiert, der aus den beiden benachbarten Fenstern bestimmt wird. Auf diese Weise berücksichtigt das erfindungsgemäße Verfahren jeweils das gesamte Sprachsignal anstatt einzel-

ner isolierter Bereiche, wodurch eine zuverlässigere Endpunktbestimmung möglich ist.

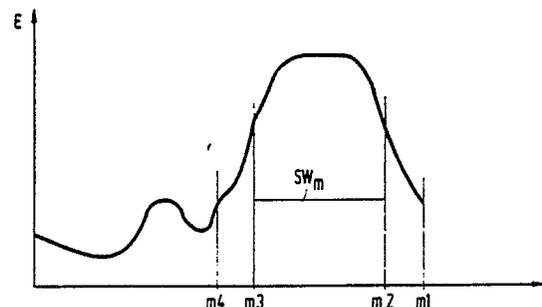


FIG. 1a

Verfahren zum Bestimmen von Anfangs- und Endpunkt isoliert gesprochener Wörter in einem Sprachsignal und Anordnung zur Durchführung des Verfahrens

Die Erfindung betrifft ein Verfahren zum Bestimmen von Anfangspunkt und Endpunkt eines Wortsignals entsprechend einem isoliert gesprochenen Wort in einem Sprachsignal durch Ermittlung eines Extremwertes in einer vom Sprachsignal abgeleiteten Folge von Digitalwerten bei Berücksichtigung von den Extremwert umgebenden Werten des Signalverlaufs und eines Schwellwertes.

Derartige Verfahren zur Bestimmung von Anfangs- und Endpunkt in einem Sprachsignal werden insbesondere verwendet, wenn das Sprachsignal aus isoliert gesprochenen Wörtern oder sehr kurzen Wortgruppen besteht und diese Wörter bzw. Wortgruppen automatisch erkannt werden sollen. Das eigentliche Wortsignal ist bei fast allen Anwendungen im Sprachsignal durch Störungen und Rauschen und Pausen sowie durch Nebengeräusche wie lautes Einatmen begleitet. Für eine möglichst zuverlässige Erkennung des oder der Wörter im Sprachsignal ist es jedoch wichtig, genau mit dem Sprachsignalteil die Erkennung zu beginnen, die auch den Anfang des zu erkennenden Wortes darstellt.

Es sind bereits verschiedene Verfahren zur Bestimmung von Anfangs- und Endpunkt bekannt. In ICASSP 84 Proceedings, 19. bis 21. März 1984, St. Diego, Kalifornien, ist auf den Seiten 18B.7.1 bis 18B.7.4 ein Verfahren zur Detektion von Endpunkten in einem Sprachsignal beschrieben, das mit der Autokorrelationsmatrix des Sprachsignals arbeitet. Eine solche Matrix ist nur mit einem erheblichen Rechenaufwand zu bilden, und die Ergebnisse sind nicht unter allen Bedingungen befriedigend. In der WO 86/03047 wird ein Endpunktdetektor verwendet, der das Sprachsignal in Blöcke unterteilt, die einander überlappen. Diese Blöcke sind jedoch festgelegt, unabhängig von dem Verlauf des Sprachsignals, und es wird der Block mit der maximalen Energie festgestellt und der davorliegende Block mit einer Energie unterhalb eines Schwellwertes bestimmt, der um ein bestimmtes Maß unter der maximalen Energie liegt. Mit weiteren aufwendigen Schritten werden dann eine Anzahl derartiger Maxima und deren Dauer ermittelt und daraus längerdauernde Energiemaxima errechnet. Auch hierbei ist insbesondere bei Überlagerung des Sprachsignals mit stärkeren Störungen eine zuverlässige Endpunkterkennung schwierig und unzuverlässig.

Aufgabe der Erfindung ist es daher, ein Verfahren der eingangs genannten Art anzugeben, das eine möglichst zuverlässige Anfangs- und Endpunktbestimmung auch bei Sprachsignalen ermög-

licht, die durch wesentliche Störsignale überlagert sind.

Diese Aufgabe wird erfindungsgemäß dadurch gelöst, daß eine Anzahl vorher aufeinanderfolgend eingetretener Digitalwerte drei benachbarten Fenstern zugeordnet werden, von denen das erste Fenster (Endfenster) eine vorgegebene erste Anzahl der zuletzt eingetretenen Digitalwerte, das zweite Fenster (Signalfenster) eine zwischen einem vorgegebenen ersten Wert und einem vorgegebenen größeren zweiten Wert variierende zweite Anzahl Digitalwerte und das dritte Fenster (Anfangsfenster) eine vorgegebene dritte Anzahl Digitalwerte umfaßt, daß für jeden neuen Digitalwert aus den Digitalwerten im ersten Fenster und aufeinanderfolgend für je einen Wert der zweiten Anzahl aus den Digitalwerten des zugehörigen dritten Fensters ein Schwellwert gebildet wird, um den jeder Digitalwert des zweiten Fensters vermindert wird, daß die Summe der so verminderten Digitalwerte für jeden Wert der zweiten Anzahl mit einer höchsten früher auf gleiche Weise gebildeten Summe verglichen und abhängig vom Vergleichsergebnis als neue höchste Summe zusammen mit Lageangaben über die Lage des zweiten Fensters innerhalb der Folge der Digitalwerte gespeichert wird, und daß die zuletzt gespeicherten Lageangaben den Anfangspunkt und den Endpunkt des Wortsignals angeben.

Es werden also keine festen Schwellwerte oder einfache absolute Maxima verwendet, sondern es werden quasi verschiedene Anfangs- und Endpunkte im Sprachsignal angenommen und dafür geprüft, ob die Energie des darin enthaltenen Sprachsignals jeweils größer ist als bei anderen angenommenen Endpunkten, wobei ein Schwellwert subtrahiert wird, der aus den benachbarten Bereichen zu beiden Seiten des angenommenen Bereichs des Wortsignals bestimmt wird. Auf diese Weise wird also kein lokales, sondern ein globales Kriterium über das gesamte Sprachsignal verwendet, indem nur ein derartiges Sprachsignal als Wortsignal gewertet wird, das über seine Umgebung maximal herausragt. Da die minimale und maximale Breite des zweiten Fensters, das also das Wortsignal darstellt, begrenzt ist, ist eine zusätzliche Sicherheit gegen Störungen gegeben und besteht außerdem die Möglichkeit, mehrere hintereinander isoliert gesprochene Wörter eindeutig voneinander zu trennen. Die Ermittlung von Anfangs- und Endpunkt erfolgt kontinuierlich mit dem Eintreffen des Sprachsignals, so daß für jede zumindest vorläufig optimale Bestimmung der Endpunkte bereits die Erkennung des Sprachsignals beginnen kann, wobei

diese abgebrochen wird, wenn ein günstigerer Wert für die Endpunkte gefunden ist, so daß auch eine schnellere Erkennung möglich ist.

Um die Zuverlässigkeit weiter zu erhöhen und beispielsweise kurze unbetonte Bereiche innerhalb eines Wortes nicht bereits als Endpunkt zu erkennen, ist es nach einer Ausgestaltung der Erfindung zweckmäßig, daß nur diejenigen Lageangaben, die für eine vorgegebene Anzahl aufeinanderfolgend eingetretener Digitalwerte unverändert geblieben sind, als Anfangspunkt und Endpunkt verwendet werden. Es wird also geprüft, ob nach dem Endpunkt eine ausreichend lange Sprachpause folgt.

Der Schwellwert, der bei der Bestimmung der Endpunkte verwendet wird, sollte möglichst weitgehend vom Störsignal ausgehen, dessen Größe jedoch nicht ohne weiteres bekannt ist. Dies erfolgt erfindungsgemäß durch Berücksichtigung je eines Bereiches vor und nach der angenommenen Lage des Wortsignals. Besonders einfach kann dieser Schwellwert dadurch gebildet werden, daß der Schwellwert aus der Summe der Digitalwerte im ersten und im dritten Fenster und eines Korrekturwertes gebildet wird. Eine derartige Summenbildung kann sehr einfach und schnell durchgeführt werden.

Als Korrekturwert kann ein fester Wert gewählt werden, der beispielsweise eine allgemeine Qualität des Sprachsignals berücksichtigt. Eine weitere Ausgestaltung der Erfindung, bei der dieser Korrekturwert den Verlauf des Sprachsignals weiter berücksichtigt, ist dadurch gekennzeichnet, daß für jeden neuen Digitalwert beim kleinsten Wert der zweiten Anzahl die Summe der Digitalwerte des zweiten Fensters gebildet und abgespeichert wird, wenn eine vorher abgespeicherte zweite Fenster-summe kleiner ist, und die Summe der Digitalwerte des dritten Fensters gebildet und abgespeichert wird, wenn eine vorher abgespeicherte dritte Fenster-summe größer ist, und aus der Differenz der beiden abgespeicherten Fenster-summen der Korrekturwert gebildet wird. Auf diese Weise gehen also nicht nur die Bereiche außerhalb der angenommenen Endpunkte ein, sondern auch das Sprachsignal zwischen den Endpunkten. Insbesondere ist es zweckmäßig, daß der Korrekturwert die durch einen konstanten vorgegebenen Signal-Rausch-Verhältniswert dividierte Differenz der beiden Fenster-summen ist. Der vorgegebene Signal-Rausch-Verhältniswert ist dann ein Maß für die mittlere Qualität des Sprachsignals und ist um so kleiner, je stärker das Sprachsignal gestört ist, wie dies beispielsweise bei Sprachübertragung über Telefonleitungen der Fall ist.

Es kann in der Praxis leicht vorkommen, daß dem Sprachsignal Störsignale überlagert sind, die zwar sehr kurz sind, jedoch eine hohe Amplitude aufweisen. Um auch in diesem Falle die Sicherheit

der Endpunkterkennung zu erhöhen, ist es nach einer weiteren Ausgestaltung der Erfindung zweckmäßig, daß als Digitalwert der kleinste von jeweils mehreren aufeinanderfolgenden digitalisierten Abtastwerten des Sprachsignals verwendet wird. Diese Maßnahme stellt ein sehr wirksames Filter für das Sprachsignal dar.

Eine Anordnung zur Durchführung des erfindungsgemäßen Verfahrens, die einen ersten Speicher zur Aufnahme von aus einem Sprachsignal abgeleiteten Digitalwerten aufweist, ist gemäß der Erfindung gekennzeichnet durch einen zweiten Speicher zur Aufnahme von Zwischenergebnissen, eine Recheneinheit, die die Digitalwerte aus dem ersten Speicher sowie Zwischenergebnisse aus dem zweiten Speicher empfängt und die Energie in jeweils einem der Fenster sowie die weiteren Zwischenergebnisse bestimmt,

einen Vergleicher zum Vergleichen von Zwischenergebnissen aus dem zweiten Speicher mit von der Recheneinheit gelieferten Werten und zum Steuern des Einschreibens der letzteren Werte in den zweiten Speicher,

eine Steuereinheit zum Adressieren des ersten und des zweiten Speichers und der Recheneinheit entsprechend den Verfahrensschritten, und

eine Zählordnung zum Zählen der unterschiedlichen zweiten Anzahlen von Digitalwerten im zweiten Fenster und zum Abgeben eines Scheleifenensignals an die Steuereinheit nach einer vorgegebenen Anzahl unterschiedlicher zweiter Anzahlen. Die Steuereinheit kann eine Ablaufsteuerung sein, die durch ein gespeichertes Programm gesteuert wird. Ein besonders einfacher Aufbau wird erhalten, wenn mindestens die Recheneinheit und die Steuereinheit durch einen Mikroprozessor realisiert sind. Dieser kann gegebenenfalls auch die Funktion des Vergleichers und der Zählordnung übernehmen.

Ausführungsbeispiele der Erfindung werden nachstehend anhand der Zeichnung erläutert. Es zeigen

Fig. 1a und 1b mit verschiedenen Lagen der Fenster,

Fig. 2a und 2b ein Flußdiagramm für den Ablauf des Verfahrens zur Endpunktbestimmung,

Fig. 3 schematisch ein Blockschaltbild einer Anordnung zur Durchführung des Verfahrens.

Der in Fig. 1a dargestellte beispielsweise Signalverlauf als Energie E oder Amplitude des Sprachsignals über die Zeit t sei momentan bis zum Zeitpunkt m_1 eingetroffen und abgetastet und liegt in Form von digitalen Abtastwerten vor. Der kontinuierlich dargestellte Signalverlauf liegt also im digitalen Bereich als Folge diskreter Punkte vor, was jedoch die weitere Erläuterung nicht wesentlich beeinflußt.

Der Signalverlauf wird nun in drei benachbarte

Fenster eingeteilt, von denen das erste Fenster von den Abtastwerten m_1 bis m_2 reicht und mit Endfenster bezeichnet wird, weil es zeitlich gesehen das vorläufige Ende des Sprachsignals darstellt. Das mittlere Fenster reicht von den Abtastwerten m_2 bis zum Abtastwert m_3 . Hierin wird das eigentliche Wortsignal angenommen, das einen höheren Energiewert als die Sprachsignalteile davor und dahinter hat. Der Punkt m_3 wird für das zu beschreibende Verfahren der Endpunktbestimmung schrittweise zwischen einem minimalen Abstand und einem maximalen Abstand vom Zeitpunkt m_2 verändert. Das dritte Fenster reicht vom jeweiligen Zeitpunkt m_3 bis zum Zeitpunkt m_4 , deren Abstand wiederum konstant ist.

Es sei bemerkt, daß jeder Abstandswert nur zu einem der Fenster gehören kann, d.h. das mittlere Fenster beginnt, wenn das erste Fenster bis zum Abtastwert zum Zeitpunkt m_2 reicht, mit dem unmittelbar links danebenliegenden Abtastwert, und entsprechendes gilt auch für das dritte Fenster. Der Einfachheit halber wird diese Tatsache in der folgenden Erläuterung nicht weiter betont, sondern weiterhin ein quasi kontinuierlicher Signalverlauf angenommen.

In Fig. 1b ist ein späterer Zeitpunkt angenommen, bei dem das Sprachsignal bereits bis zum Zeitpunkt n_1 eingetroffen ist. Außerdem ist ein größeres Signalfenster angenommen, so daß dessen Beginn beim Zeitpunkt n_3 weiter vom Zeitpunkt n_2 entfernt liegt als bei Fig. 1a. Entsprechend liegt dann auch der Zeitpunkt n_4 als Beginn des Anfangsfensters zu einem noch früheren Zeitpunkt.

Ein wesentliches Kriterium bei der Bestimmung der Endpunkte des Sprachsignals ist die Fläche des Sprachsignals innerhalb des Signalfensters, vermindert um einen Schwellwert SW , der u.a. von der Fläche unter dem Sprachsignal im ersten und im dritten Fenster abhängt. Die Flächen unter dem Sprachsignal werden im übrigen durch die Summe der digitalisierten Abtastwerte innerhalb des jeweiligen Fensters dargestellt.

In Fig. 1a ist die Fläche in dem Anfangs- und Endfenster noch relativ groß, so daß sich ein höherer Schwellwert SW_m ergibt. Aus der Figur ist unmittelbar zu erkennen, daß die um den Schwellwert verminderte Fläche im mittleren Fenster größer wird, wenn Anfangs- und Endfenster weiter auseinandergezogen werden, d.h. wenn die noch folgenden eintreffenden Teile des Signalverlaufs abgewartet werden und die Breite des Signalfensters größer gewählt wird.

In Fig. 1b ist dieser Fall nun dargestellt, wobei die Fläche unter dem Sprachsignal im Anfangs- und im Endfenster nun wesentlich kleiner ist, so daß auch der Schwellwert SW_n niedriger liegt, jedoch ist nun zu erkennen, daß die Teile des Sprachsignals, die dem Anfangs- und Endfenster

am nächsten liegen, einen negativen Beitrag zur Gesamtfläche im Signalfenster abzüglich des Schwellwertes SW_n liefern, da diese Signalwerte kleiner sind als der Schwellwert. Bei der optimalen Entscheidung fallen Anfangs- und Endpunkt auf Zeiten, bei denen der Signalwert gleich dem Schwellwert ist. Der Bereich des Sprachsignals, der innerhalb dieses Signalfensters kurzzeitig unterhalb des Schwellwertes SW_n liegt, liefert zwar ebenfalls einen negativen Beitrag, der jedoch durch den links davon liegenden höheren Signalabschnitt übertroffen wird, so daß sich durch Verlängern des mittleren Fensters über diesen Bereich des Sprachsignals hinaus insgesamt eine Vergrößerung der Gesamtfläche im Signalfenster oberhalb des Schwellwertes SW_n ergibt. Der vorher genannte Anfangs- und Endpunkt wird mit dem Verfahren entsprechend dem Flußdiagramm in Fig. 2a und 2b bestimmt.

Das Symbol 10 bedeutet den Start des gesamten Ablaufs, d.h. den Beginn des Sprachsignals. Im Block 11 werden verschiedene Anfangswerte eingestellt, eine Anzahl Abtastwerte entsprechend der Länge des Endfensters, des minimalen Signalfensters und des Anfangsfensters abgewartet, bevor das Verfahren starten kann, sowie eine spezielle Filterfunktion durchgeführt. Diese besteht darin, daß von jeweils drei aufeinanderfolgenden Abtastwerten der kleinste Wert gewählt und als Digitalwert dem Verfahren zugeführt wird. Beispielsweise wird alle 10 ms ein Abtastwert dem Sprachsignal entnommen, der den Momentanwert oder den integrierten Wert seit dem letzten Abtastwert darstellt, und die Abtastwerte werden digitalisiert. Wenn von jeweils drei aufeinanderfolgenden Abtastwerten der kleinste ausgewählt wird, werden dem Verfahren also alle 30 ms ein Digitalwert zugeführt, so daß für die Durchführung der folgenden Verfahrensschritte 30 ms zur Verfügung stehen. Die zugeführten Digitalwerte werden gespeichert, da sie zu späteren Zeitpunkten noch benötigt werden, und zwar mindestens über eine Signaldauer, die der Summe aus der vorgegebenen maximalen Dauer des Signalfensters und der beiden anderen Fenster entspricht.

Im Block 12 wird die Energie EF_k im Anfangsfenster zwischen den Punkten m_3 und m_4 in Fig. 1a bzw. n_3 und n_4 in Fig. 1b durch Aufsummieren der darin liegenden Signalwerte bestimmt. Im Block 13 wird dieser Wert durch die Länge B_F des Anfangsfensters dividiert und damit die mittlere Energie eF_k in diesem Fenster bestimmt.

In einem Vergleich 14 wird geprüft, ob dieser mittlere Wert eF_k kleiner ist als ein gespeicherter Wert eF_{sp} , und wenn dies der Fall ist, wird im Block 15 dieser kleinere Wert gespeichert, d.h. eF_{sp} wird durch den momentanen Wert eF_k ersetzt. Nach dem Block 15 oder wenn der neue Wert nicht kleiner ist als der gespeicherte Wert wird im Block

16 die Energie ES_k des Signalfensters mit minimaler Länge bestimmt, also die Fläche unter dem Sprachsignalverlauf zwischen den Punkten m_2 und m_3 in Fig. 1a, wofür ebenfalls die gespeicherten Digitalwerte in diesem Bereich aufsummiert werden. Danach wird in einem Vergleich 17 geprüft, ob diese Energie ES_k größer ist als eine gespeicherte Energie ES_{sp} . Wenn dies der Fall ist, wird im Block 18 der gespeicherte Wert durch den neuen Wert ersetzt, danach oder wenn der neue Wert nicht größer ist als der gespeicherte Wert, wird im Block 20 die mittlere Energie ES_k bestimmt, indem die Gesamtenergie ES_k durch die minimale Breite B_{s0} des Signalfensters dividiert wird. Die Breite B dieses Fensters sowie der anderen Fenster wird jeweils durch die Anzahl darinliegender Digitalwerte angegeben.

Danach wird im Block 21 ein Korrekturwert thN aus der Differenz zwischen der mittleren Energie eS_k im Signalfenster und eF_k im Anfangsfenster, die durch einen angenommenen Signal-Rausch-Verhältniswert SNR dividiert wird, bestimmt. Schließlich wird im Block 22 die mittlere Energie im Endfenster, also zwischen den Punkten m_1 und m_2 in Fig. 1a oder n_1 und n_2 in Fig. 1b auf entsprechende Weise wie im Anfangsfenster bestimmt.

Die Schritte 12 bis 22 laufen bei jedem neu eingetroffenen Digitalwert einmal ab, während der Verbindungspunkt 23 nun zu einer Schleife führt, die für jede zugelassene Breite des Signalfensters einmal durchlaufen wird. Diese einzelnen Durchläufe werden mit dem Index 1 angedeutet.

In Fig. 2b ist diese Schleife angegeben, die mit dem Verbindungspunkt 23 beginnt. Im Block 29 wird dieser Wert 1 auf den Anfangswert Null gesetzt. Im folgenden Block 30 wird der mittlere Energiewert eF_1 des Anfangsfensters bei der jeweiligen Verschiebung 1 von der minimalen Breite des Signalfensters aus entsprechend Block 13 bestimmt, und im Block 31 wird der so ermittelte Wert zu dem im Block 22 ermittelten mittleren Energiewert des Endfensters und zu dem im Block 21 ermittelten Korrekturwert thN addiert, um den Schwellwert thr zu ermitteln. Danach wird im Block 32 die Energie ES_1 des Signalfensters in der jeweiligen Breite durch Aufsummieren der Digitalwerte in diesem Fenster bestimmt. Im Block 33 wird schließlich von diesem Wert der Schwellwert thr , multipliziert mit der jeweiligen Breite B_{s1} des Signalfensters, subtrahiert. Dies ist die Fläche unter dem Signalverlauf in Fig. 1a zwischen den Punkten m_2 und m_3 bzw. in Fig. 1b zwischen den Punkten n_2 und n_3 , vermindert um die Fläche unterhalb des Schwellwertes SW_m bzw. SW_n zwischen diesen Punkten. Diese effektive Energie EPS_1 wird als die Energie des Sprachsignals in dem Signalfenster angesehen, die über das Störsignal hinausragt, wo-

bei dieses Störsignal nicht direkt ermittelt werden kann, sondern ein wahrscheinlicher Wert in Form des Schwellwertes in der vorher beschriebenen Weise abgeleitet wird.

5 Beim Vergleich 34 wird geprüft, ob diese zuletzt ermittelte effektive Energie EPS_e des Sprachsignals über einem gespeicherten Wert EPS_s liegt. Wenn dies der Fall ist, wird im Block 35 dieser neue Wert abgespeichert, ferner wird gespeichert, bei welchem zuletzt eingetroffenen Digitalwert dies erfolgt ist, indem ein momentaner Index k als Wert k_{sp} gespeichert wird, und ferner werden Anfangs- und Endpunkt des Signalfensters, d.h. die Werte m_2 und m_3 in Fig. 1a bzw. n_2 und n_3 in Fig. 1b gespeichert. Daran anschließend oder, wenn beim Vergleich 34 der neue Wert nicht größer ist als der gespeicherte, wird im Block 36 der Schleifenwert 1 um 1 erhöht, und im Vergleich 37 wird geprüft, ob dieser Wert 1 den vorgegebenen Maximalwert L entsprechend der maximalen Breite des Signalfensters erreicht hat. Wenn dies nicht der Fall ist, wird zum Block 30 zurückgegangen.

Anderenfalls wird im Vergleich 38 geprüft, ob das gefundene Maximum der Energie im Sprachfenster stationär ist, d.h. ob eine ausreichende Anzahl K_{ST} von weiteren Digitalwerten zugeführt worden ist, ohne daß ein höherer Energiewert gefunden worden ist. Wenn dies nicht der Fall ist, wird zum Block 12 zurückgegangen und der nächste Digitalwert verarbeitet. Wenn jedoch seit einer vorgegebenen Anzahl neu zugeführter Digitalwerte keine höhere Energie im Signalfenster festgestellt worden ist, wird angenommen, daß die zuletzt im Block 35 abgespeicherte effektive Energie dasjenige Signalfenster angibt, das dem Wortsignal innerhalb des Sprachsignals am besten entspricht, und die dabei abgespeicherten Lagewerte des Fensters, d.h. die Punkte m_2 und m_3 bzw. n_2 und n_3 geben den gesuchten Anfangspunkt und Endpunkt des Wortsignals an.

In dem Flußdiagramm in Fig. 2a und 2b sind nur die notwendigsten Verfahrensschritte angegeben. Es sind Einsparungen bei der Durchführung der Rechenschritte in dem Verfahren insbesondere dadurch möglich, daß Zwischenwerte gespeichert werden. Beispielsweise können die Energiewerte EF_k bzw. die entsprechenden mittleren Energiewerte, die in dem Block 12 bzw. 13 ermittelt worden sind, stets zwischengespeichert werden, da diese bei den folgenden zugeführten Digitalwerten wieder verwendet werden können, denn das Anfangsfenster für die kleinste Breite des Signalfensters für einen bestimmten Digitalwert hat die gleiche Lage wie das Anfangsfenster bei dem folgenden Digitalwert, wenn das Signalfenster um eine Einheit gegenüber dem minimalen Wert vergrößert ist, usw. Entsprechendes gilt auch für die Energie im Signalfenster. Diese Einsparung an Rechenzeit erfor-

dert jedoch einen höheren Speicherungs- und Adressensteuerungsaufwand für die Zwischenspeicher.

Wenn das beschriebene Verfahren in Verbindung mit automatischer Spracherkennung verwendet wird, kann der Erkennungsvorgang bei jedem neuen Abspeichern der Werte im Block 35 beginnen, so daß dann, wenn der stationäre Zustand im Block 38 schließlich erkannt ist, das Erkennungsverfahren bereits weit fortgeschritten sein kann, so daß auf diese Weise eine schnellere Erkennung, ggf. eine Erkennung in Echtzeit, möglich ist.

Bei der in Fig. 3 dargestellten Anordnung nimmt ein Schallwandler 40 ein Sprachsignal auf und setzt es in ein elektrisches Signal um. Dieses wird einer Einheit 42 zugeführt, die dem kontinuierlichen Signal in regelmäßigen Zeitabständen Abtastwerte entnimmt und diese digitalisiert. Die Einheit 44 wählt aus jeweils drei aufeinanderfolgenden digitalisierten Abtastwerten den kleinsten aus und führt die so ermittelten Digitalwerte einem Speicher 50 zu. Wenn die Einheit 42 alle 10 ms einen Abtastwert dem Sprachsignal entnimmt, erhält der Speicher 50 also alle 30 ms einen neuen Digitalwert. Dieser wird an einer Adresse abgespeichert, die von einer Steuereinheit 52 über die Verbindung 53 zugeführt wird.

Entsprechend adressiert die Steuereinheit 52 den Speicher 50 auch zum Auslesen der gespeicherten Digitalwerte, die einer Recheneinheit 54 zugeführt werden. Diese wird ebenfalls von der Steuereinheit 52 über eine Verbindung 51 gesteuert und führt die Rechenschritte aus, die in dem Flußdiagramm in Fig. 2a und 2b mit den Blöcken 12, 13, 16, 20 bis 22 und 30 bis 33 angegeben sind. Insbesondere bestimmt die Recheneinheit 54 die Energie im Anfangsfenster durch Aufsummieren der entsprechenden, von der Steuereinheit im Speicher 50 adressierten Digitalwerte und bildet die mittlere Energie. Diese wird über die Leitung 55 einem Vergleichler 58 zugeführt, der am anderen Eingang aus einem zweiten Speicher 56 über dessen Datenausgangsleitung 57 den entsprechenden früher gespeicherten Wert empfängt. Der zweite Speicher 56 wird dabei ebenfalls von der Steuereinheit 52 über die Leitung 59 adressiert. Wenn der auf der Leitung 55 vorhandene, neu ermittelte Wert kleiner ist als der auf der Leitung 57 vorhandene gespeicherte Wert, erzeugt der Vergleichler 58 ein entsprechendes Signal und führt es dem zweiten Speicher 56 zu, so daß nun an der adressierten Stelle der neue, auf der Leitung 55 vorhandene Wert gespeichert wird. Dies entspricht den Blöcken 14 und 17 in Fig. 2a. In entsprechender Weise werden auch die anderen Berechnungen und Vergleiche durchgeführt, wobei die Recheneinheit 54 insbesondere bei den Schritten 21, 31 und 33 die dort erforderlichen Werte aus dem zweiten

Speicher 56 über die Leitung 57 erhält. Für das Abspeichern der weiteren Werte beim Schritt 35 führt die Steuereinheit 52 diese Werte über die Leitung 69 dem Dateneingang des zweiten Speichers 56 zu.

Ferner ist noch ein Zähler 60 vorhanden, der den Index l zählt. Über die Leitung 65 wird der Zähler 60 von der Steuereinheit 52 auf die Anfangsstellung gesetzt und mit Zähltaktten versorgt, wie bei den Schritten 29 und 36 in Fig. 2b angegeben ist. Jeweils wenn der Zähler 60 eine Anzahl L Taktsignale empfangen hat, die dem Unterschied zwischen dem kleinsten und dem größten Signalfenster entspricht, gibt er über die Leitung 63 ein Schleifenendesignal an die Steuereinheit 52 ab. Dies entspricht dem Vergleich 37 in Fig. 2b. Der Vergleich 38 wird zweckmäßig in der Steuereinheit 52 durchgeführt.

Eine einfache Realisierung der Anordnung nach Fig. 3 besteht darin, daß die Steuereinheit 52 und die Recheneinheit 54 durch einen Mikroprozessor gebildet werden. Dieser kann dann auch noch die Funktion des Vergleichlers 58 und des Zählers 60 übernehmen, so daß sich insgesamt ein sehr einfacher Aufbau ergibt.

Ansprüche

1. Verfahren zum Bestimmen von Anfangspunkt und Endpunkt eines Wortsignals entsprechend einem isoliert gesprochenen Wort in einem Sprachsignal durch Ermittlung eines Extremwertes in einer vom Sprachsignal abgeleiteten Folge von Digitalwerten bei Berücksichtigung von den Extremwert umgebenden Werten des Signalverlaufs und eines Schwellwertes, dadurch gekennzeichnet, daß eine Anzahl vorher aufeinanderfolgend eingetretener Digitalwerte drei benachbarten Fenstern zugeordnet werden, von denen das erste Fenster (Endfenster) eine vorgegebene erste Anzahl (B_R) der zuletzt eingetretenen Digitalwerte, das zweite Fenster (Signalfenster) eine zwischen einem vorgegebenen ersten Wert und einem vorgegebenen größeren zweiten Wert variierende zweite Anzahl (B_{S1}) Digitalwerte und das dritte Fenster (Anfangsfenster) eine vorgegebene dritte Anzahl (B_F) Digitalwerte umfaßt, daß für jeden neuen Digitalwert aus den Digitalwerten im ersten Fenster und aufeinanderfolgend für je einen Wert (1) der zweiten Anzahl (B_{S1}) aus den Digitalwerten des zugehörigen dritten Fensters ein Schwellwert (thr) gebildet wird, um den jeder Digitalwert des zweiten Fensters vermindert wird, daß die Summe der so verminderten Digitalwerte für jeden Wert der zweiten Anzahl mit einer höchsten früher auf gleiche Weise gebildeten Summe verglichen und abhängig vom Vergleichsergebnis als

neue höchste Summe zusammen mit Lageangaben über die Lage des zweiten Fensters innerhalb der Folge der Digitalwerte gespeichert wird, und daß die zuletzt gespeicherten Lageangaben den Anfangspunkt und den Endpunkt des Wortsignals angeben.

2. Verfahren nach Anspruch 1,

dadurch gekennzeichnet, daß nur diejenigen Lageangaben, die für eine vorgegebene Anzahl aufeinanderfolgend eingetretener Digitalwerte unverändert geblieben sind, als Anfangspunkt und Endpunkt verwendet werden.

3. Verfahren nach Anspruch 1 oder 2,

dadurch gekennzeichnet, daß der Schwellwert aus der Summe der Digitalwerte im ersten und im dritten Fenster und eines Korrekturwertes gebildet wird.

4. Verfahren nach Anspruch 3,

dadurch gekennzeichnet, daß für jeden neuen Digitalwert beim kleinsten Wert der zweiten Anzahl (B_{S0}) die Summe der Digitalwerte des zweiten Fensters gebildet und abgespeichert wird, wenn eine vorher abgespeicherte zweite Fenstersumme kleiner ist, und die Summe der Digitalwerte des dritten Fensters gebildet und abgespeichert wird, wenn eine vorher abgespeicherte dritte Fenstersumme größer ist, und aus der Differenz der beiden abgespeicherten Fenstersummen der Korrekturwert gebildet wird.

5. Verfahren nach Anspruch 4,

dadurch gekennzeichnet, daß der Korrekturwert die durch einen konstanten vorgegebenen Signal-Rausch-Verhältniswert dividierte Differenz der beiden Fenstersummen ist.

6. Verfahren nach einem der Ansprüche 1 bis 5,

dadurch gekennzeichnet, daß als Digitalwert der kleinste von jeweils drei aufeinanderfolgenden digitalisierten Abtastwerten des Sprachsignals verwendet wird.

7. Anordnung zur Durchführung des Verfahrens nach einem der Ansprüche 1 bis 6, mit einem ersten Speicher zum Aufnehmen von aus einem Sprachsignal abgeleiteten Digitalwerten,

gekennzeichnet durch einen zweiten Speicher zur Aufnahme von Zwischenergebnissen,

eine Recheneinheit, die die Digitalwerte aus dem ersten Speicher sowie Zwischenergebnisse aus dem zweiten Speicher empfängt und die Energie in jeweils einem der Fenster sowie die weiteren Zwischenergebnisse bestimmt,

einen Vergleich zum Vergleichen von Zwischenergebnissen aus dem zweiten Speicher mit von der Recheneinheit gelieferten Werten und zum Steuern des Einschreibens der letzteren Werte in den zweiten Speicher,

eine Steuereinheit zum Adressieren des ersten und

des zweiten Speichers und der Recheneinheit entsprechend den Verfahrensschritten, und eine Zählordnung zum Zählen der unterschiedlichen zweiten Anzahlen von Digitalwerten im zweiten Fenster und zum Abgeben eines Schleifenendesignals an die Steuereinheit nach einer vorgegebenen Anzahl unterschiedlicher zweiter Anzahlen.

8. Anordnung nach Anspruch 7,

dadurch gekennzeichnet, daß mindestens die Recheneinheit und die Steuereinheit durch einen Mikroprozessor realisiert sind.

5

10

15

20

25

30

35

40

45

50

55

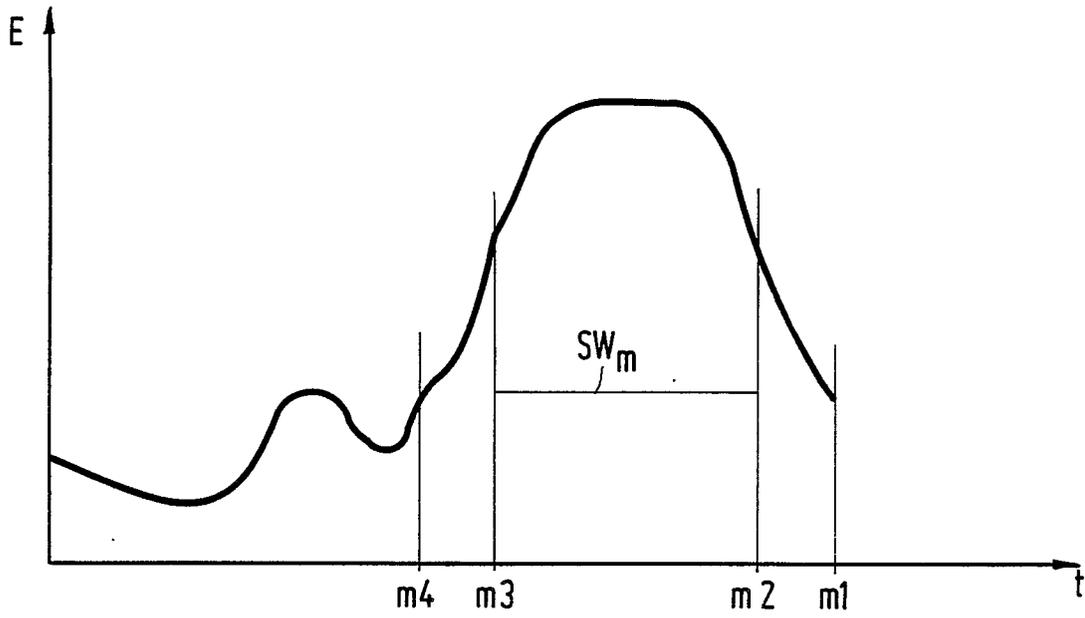


FIG. 1a

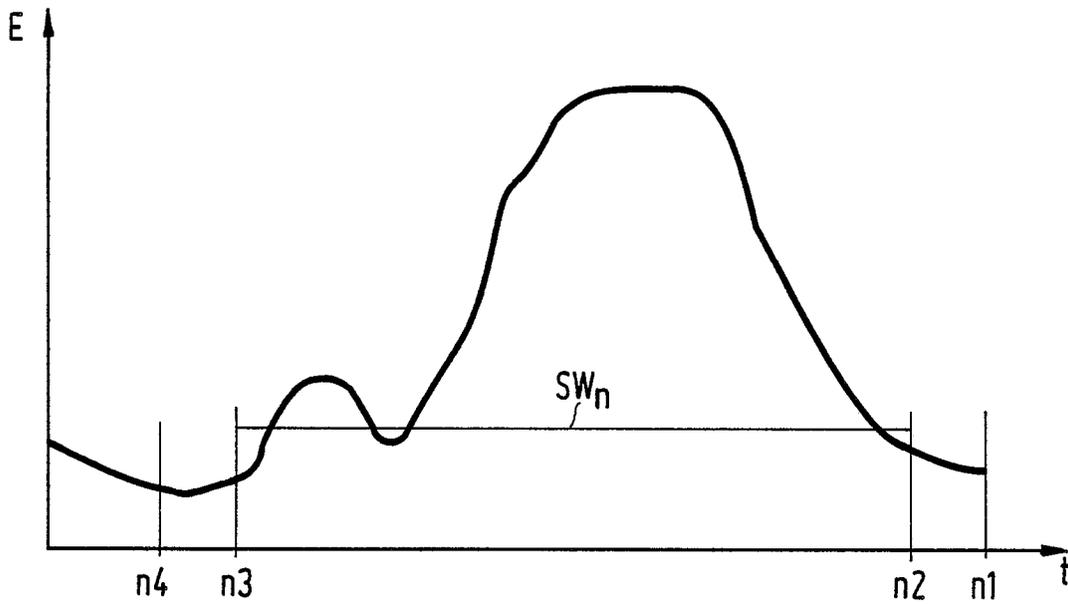


FIG. 1b

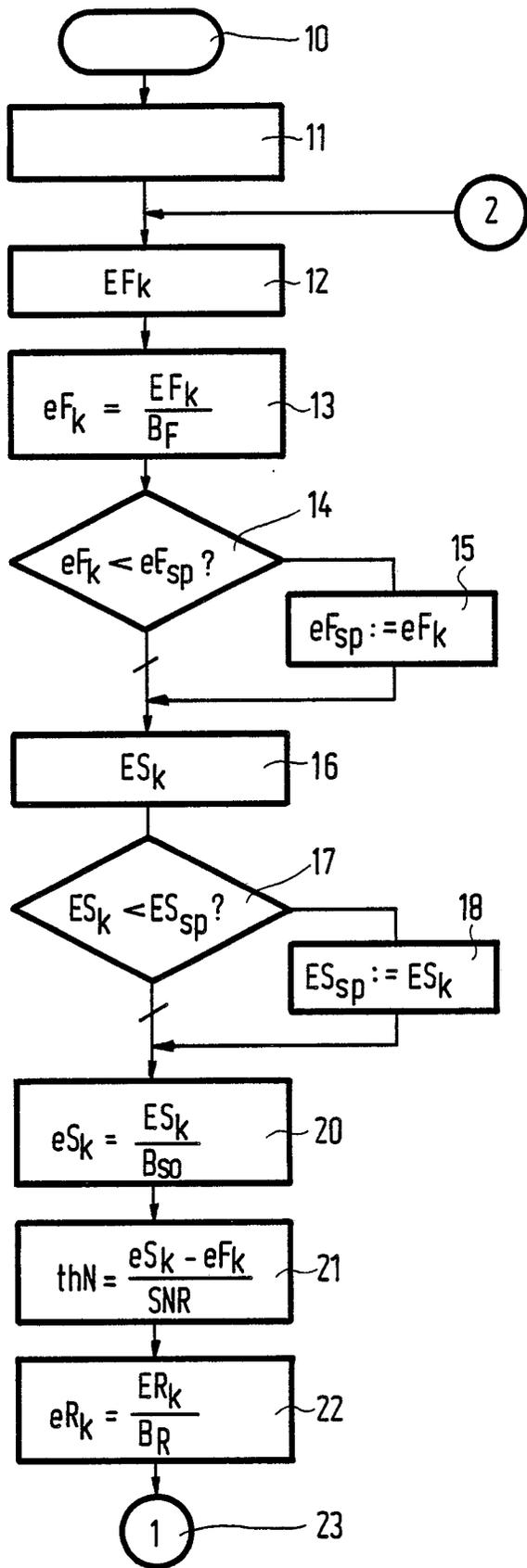


FIG. 2a

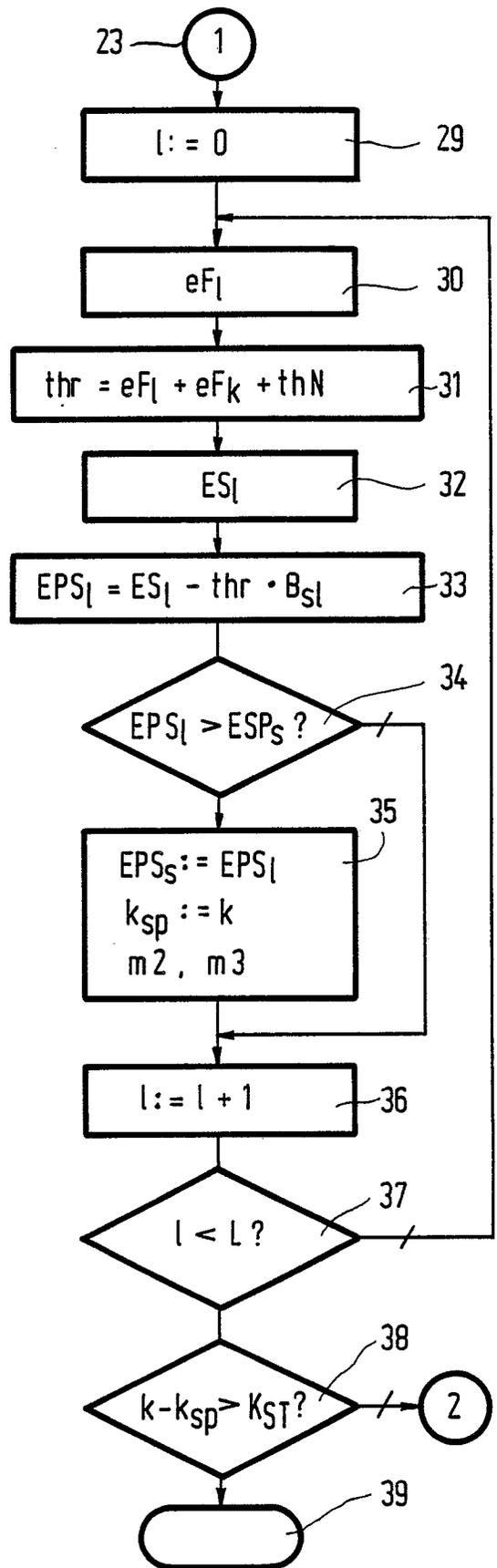


FIG. 2b

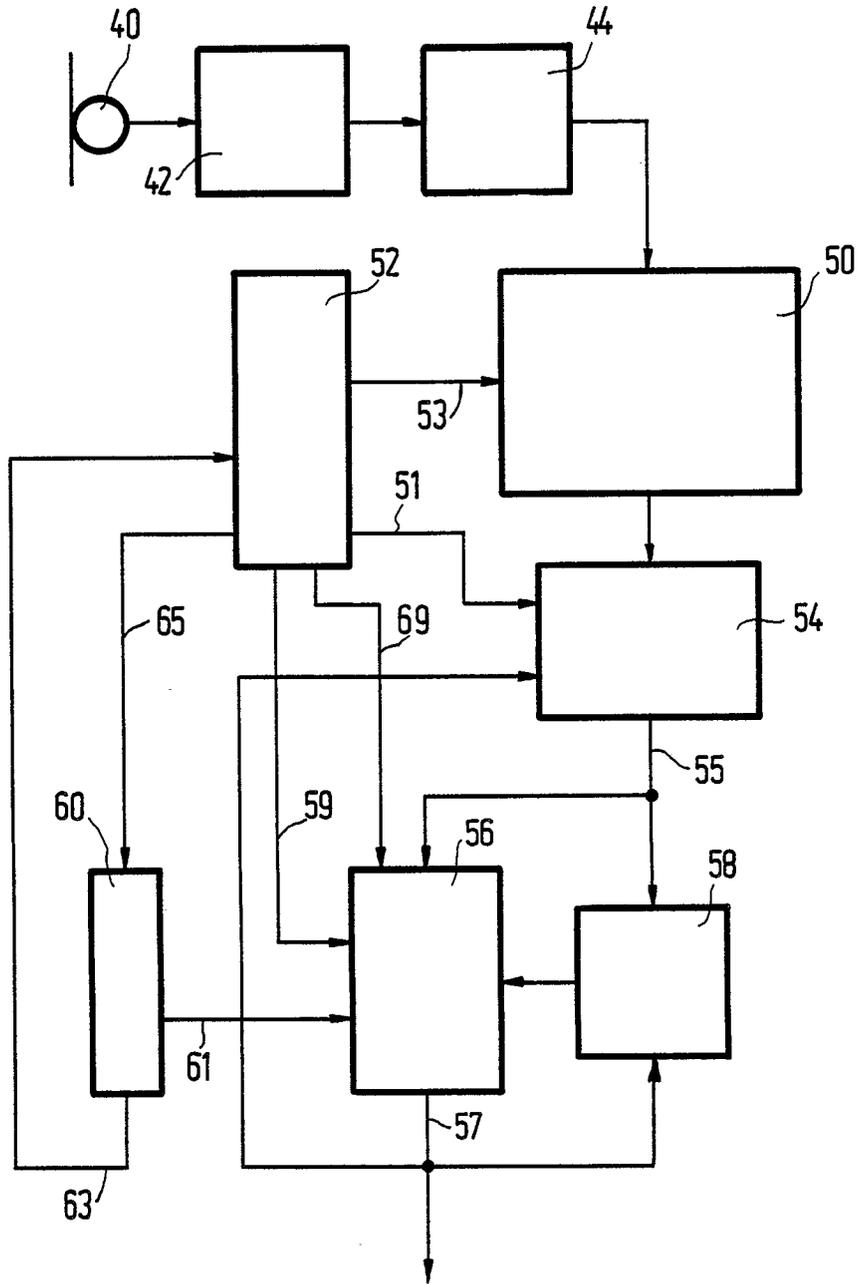


FIG. 3