



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 601 20 062 T2** 2006.11.16

(12)

Übersetzung der europäischen Patentschrift

(97) **EP 1 189 206 B1**

(21) Deutsches Aktenzeichen: **601 20 062.4**

(96) Europäisches Aktenzeichen: **01 402 126.5**

(96) Europäischer Anmeldetag: **06.08.2001**

(97) Erstveröffentlichung durch das EPA: **20.03.2002**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **31.05.2006**

(47) Veröffentlichungstag im Patentblatt: **16.11.2006**

(51) Int Cl.⁸: **G10L 15/26** (2006.01)

G10L 17/00 (2006.01)

G06F 3/16 (2006.01)

(30) Unionspriorität:

10046561 **19.09.2000** **DE**

(73) Patentinhaber:

Thomson Licensing, Boulogne Billancourt, FR

(74) Vertreter:

**Rittner, K., Dipl.-Phys. Dr.rer.nat., Pat.-Ass., 30453
Hannover**

(84) Benannte Vertragsstaaten:

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

(72) Erfinder:

**Schröder, Ernst F., 30655 Hannover, DE; Platte,
Hans-Joachim, 30966 Hemmingen, DE**

(54) Bezeichnung: **Sprachsteuerung von elektronischen Geräten**

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] Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur Sprachsteuerung von einem Gerät der Unterhaltungselektronik

Stand der Technik

[0002] Die Bedienung von Unterhaltungselektronikgeräten, wie z.B. Fernsehgeräten oder Videorekordern, kann für den Benutzer durch eine Sprachsteuerung vereinfacht werden. So ist es bekannt, sprachgesteuert Geräteeinstellungen zu verändern, Bedienfunktionen wie z.B. eine Senderwahl auszuführen oder Programmierungen vorzunehmen.

[0003] Die von dem Benutzer gesprochenen Bedienbefehle werden hierfür zunächst als Schallsignale detektiert, in elektrische Signale gewandelt und digitalisiert. Die digitalisierten Sprachsignale werden dann einem Spracherkennungssystem zugeführt. Die Spracherkennung basiert hierbei üblicherweise auf einem Akustik- und einem Sprachmodell. Das akustische Modell nutzt eine große Anzahl von Sprachmustern, wobei mathematische Algorithmen dazu verwendet werden, die akustisch am besten passenden Worte zu einem gesprochenen Wort anzugeben. Das Sprachmodell wiederum basiert auf einer Analyse, bei der anhand von einer Vielzahl von Dokumentproben festgestellt wird, in welchem Kontext und wie häufig gewisse Wörter normalerweise verwendet werden.

[0004] Gegenwärtige Systeme sehen vor, die Bedienbefehle in das in einer Fernbedienung integrierte Mikrofon zu sprechen. Indem die Fernbedienung unmittelbar vor den Mund des Benutzers gehalten wird, wird so eine Verschlechterung der Erkennungsrate durch störende Hintergrundgeräusche verhindert. Dieses erfordert jedoch, wie bei herkömmlichen Fernbedienungen auch, daß der Benutzer die Fernbedienung weiterhin zur Hand nehmen muß. Ein Komfortgewinn läßt sich erreichen, wenn für die Spracheingabe ein oder mehrere Mikrofone in dem Unterhaltungselektronikgerät vorgesehen sind, so daß der Benutzer eine Bedienung von einer beliebigen Stelle des Raumes ohne Mitführen der Fernbedienung vornehmen kann. Die erforderliche Unterdrückung von Hintergrundstörungen kann in diesem Fall durch die Verwendung spezieller Mikrofon-Arrays und Verfahren wie dem sogenannten „Statistical beamforming“ oder „blind source Separation“ erfolgen. Das bediente Gerät ist aber nicht in der Lage zu ermitteln, welche Spracheingaben von dem gegenwärtigen Benutzer erfolgen. Es ist daher nicht möglich, und nur auf diese Bedienbefehle zu reagieren, Äußerungen von anderen Personen dagegen zu ignorieren.

[0005] Ein weiterer Ansatz zur Erhöhung des Benut-

zerkomforts wird in der automatischen Zwischenspeicherung von Fernsehprogrammen auf in Fernsehern oder Set-Top-Boxen integrierten Festplatten gesehen. Nach einer Analyse der Sehgewohnheiten werden hierbei automatisch die Programme oder Programmarten aufgezeichnet, die der Benutzer vorher regelmäßig gewählt hat. Wenn der Benutzer dann zu irgendeiner Zeit seinen Fernseher einschaltet, kann er mit gewisser Wahrscheinlichkeit seine Lieblingssendungen anschauen. Die Qualität der Analyse wird jedoch im Falle mehrerer Benutzer dadurch beeinträchtigt, daß nicht unterschieden werden kann, welcher Benutzer den Fernseher wann bedient.

[0006] WO 00/39789 beschreibt die Kombination von Spracherkennungs- und Sprecherverifikations-Techniken um nur eine Ausführung von durch autorisierte Benutzer geäußerte Spracheingaben, zu garantieren.

Erfindung

[0007] Der Erfindung liegt die Aufgabe zugrunde, ein Verfahren zur Sprachsteuerung anzugeben, welches die vorgenannten Nachteile vermeidet. Diese Aufgabe wird durch das in Anspruch 1 angegebene Verfahren gelöst.

[0008] Im Prinzip besteht das Verfahren zur Sprachsteuerung von einem Gerät der Unterhaltungselektronik darin, daß Spracheingaben eines Benutzers in digitalisierte Sprachsignale umgewandelt werden. Aus den digitalisierten Sprachsignalen werden erste Merkmale extrahiert, die charakteristisch für die einzelnen Laute der Sprache sind und so eine Erkennung der gesprochenen Laute ermöglichen. Weiterhin werden aus den digitalisierten Sprachsignalen zweite Merkmale extrahiert, die eine Charakterisierung der Stimme des jeweiligen Benutzers ermöglichen und für eine Unterscheidung der Spracheingaben von verschiedenen Benutzern verwendet werden. Nach einem Sprachkommando von einem ersten Benutzer werden weitere Sprachkommandos nur noch von diesem ersten Benutzer akzeptiert, indem die weiteren Spracheingaben auf charakteristische Sprachmerkmale geprüft werden und nur dann akzeptiert werden, wenn sie aufgrund dieser Merkmale demselben Sprecher zuzuordnen sind.

[0009] Hiermit kann sichergestellt werden, daß zeitabschnittsweise nur einer von mehreren gleichzeitigen Benutzern das betreffende Gerät per Sprache bedienen kann – ähnlich dem Fall, daß nur einer von mehreren Benutzern über eine passende Fernbedienung verfügt.

[0010] Insbesondere kann es von Vorteil sein, daß ein Sprachkommando zum Einschalten des Gerätes von einem beliebigen ersten Benutzer akzeptiert wird und danach Sprachkommandoeingaben nur noch

von diesem ersten Benutzer akzeptiert werden.

[0011] Ein Sprachkommando zum Ausschalten des Gerätes kann vorzugsweise nur von dem ersten Benutzer akzeptiert werden, wobei nach dem Ausschalten des Gerätes wieder Sprachkommandos von einem beliebigen Benutzer akzeptiert werden.

[0012] Für bestimmte Anwendungen kann es jedoch auch von Vorteil sein, daß ein Sprachkommando zum Ausschalten des Gerätes von einem beliebigen Benutzer akzeptiert wird.

[0013] Ebenso kann vorteilhafterweise ein Bedienbefehl vorgesehen sein, ab dessen Eingabe durch den ersten Benutzer Sprachkommandos von einem zweiten Benutzer akzeptiert werden. Dieses ermöglicht, die Bediengewalt weiterzugeben, entsprechend dem Weiterreichen einer Fernbedienung von einem ersten zu einem zweiten Benutzer.

[0014] Besonders vorteilhaft kann es sein, daß eine Identifizierung der verschiedenen Benutzer erfolgt, um eine Analyse der Sehgewohnheiten vorzunehmen und hieraus Benutzerprofile der verschiedenen Benutzer zu erstellen.

[0015] Vorzugsweise wird ein so gewonnenes Benutzerprofil bei einer Zwischenspeicherung von Fernsehprogrammen verwendet, um eine getrennte Zwischenspeicherung von bevorzugten Sendungen für verschiedene Benutzer zu ermöglichen.

[0016] Ebenso kann das Benutzerprofil verwendet werden, um Vorschläge für zu betrachtende Sendungen vorzunehmen, die auf den Sehgewohnheiten des verschiedenen Benutzers abgestimmt sind.

Zeichnungen

[0017] Anhand der Zeichnungen werden Ausführungsbeispiele der Erfindung beschrieben. Diese zeigen in

[0018] [Fig. 1](#) schematisch den Verfahrensablauf zur Unterscheidung der Sprachkommandos von verschiedenen Benutzern, um zukünftige Sprachkommandos nur noch von bestimmten Benutzern zu akzeptieren;

[0019] [Fig. 2](#) schematisch den Verfahrensablauf zur Identifizierung verschiedener Benutzer, um hieraus Benutzerprofile der verschiedenen Benutzer zu erstellen.

Ausführungsbeispiele

[0020] In [Fig. 1](#) ist schematisch der Ablauf eines ersten Ausführungsbeispiels in Form eines Flußdiagramms dargestellt. Bei einem sprachgesteuerten

Fernseher erfolgt hier eine Unterscheidung der Spracheingaben von verschiedenen Benutzern, um zukünftige Spracheingaben nur noch von bestimmten Benutzern zu akzeptieren.

[0021] Zunächst werden in einem ersten Verfahrensschritt **1** die Schallsignale in elektrische Signale umgewandelt, so daß ein analoges Sprachsignal vorliegt, welches wiederum in ein digitales Sprachsignal umgewandelt wird.

[0022] Dann werden in einem nächsten Verfahrensschritt **2** aus dem digitalisierten akustischen Signal erste Merkmale gewonnen, die möglichst typisch für die einzelnen Laute der Sprache und robust gegenüber Störungen und Schwankungen in der Aussprache sind. Ebenso werden in dem Verfahrensschritt **3** aus dem digitalisierten akustischen Signal zweite Merkmale extrahiert, die eine Charakterisierung der Stimme des jeweiligen Benutzers ermöglichen und für eine Unterscheidung der Spracheingaben von verschiedenen Benutzern verwendet werden. Diese Merkmalsextraktion erfolgt in dem dargestellten Ausführungsbeispiel getrennt für den Spracherkennung und den Sprechererkennung, kann aber auch gemeinsam erfolgen.

[0023] Basierend auf den ersten Merkmalen erfolgt dann die eigentliche Spracherkennung im Verfahrensschritt **4**. In Verfahrensschritt **5** wird mit Hilfe der zweiten Merkmale eine Sprechererkennung durchgeführt, um den gerade Sprechenden Benutzer zu identifizieren. Ebenso kann jedoch lediglich eine Abspeicherung der zweiten Merkmale erfolgen, um so eine Unterscheidung zu anderen Benutzern zu ermöglichen ohne das eine Identifizierung der einzelnen Benutzer erfolgt.

[0024] In dem Verfahrensschritt **6** wird dann geprüft, ob der Fernseher bereits eingeschaltet ist. Wenn dieses der Fall ist, werden die Verfahrensschritte **7** und **8** durchlaufen, anderenfalls die Verfahrensschritte **9** und **10**. Sollte der Fernseher noch nicht eingeschaltet sein, so wird als nächstes im Verfahrensschritt **9** geprüft, ob ein Einschaltkommando wie beispielsweise „An“, „Fernseher an“ vorliegt. Falls dieses der Fall ist, so wird im Verfahrensschritt **10** der Fernseher eingeschaltet und der Benutzer, von dem die Eingabe stammt, vermerkt. Falls statt einer Identifizierung lediglich eine Unterscheidung von verschiedenen Benutzern erfolgt, werden entsprechend die zweiten Merkmale, die den aktuellen Benutzer charakterisieren, abgespeichert. Anschließend wird, ebenso wie für den Fall, daß im Verfahrensschritt **9** kein Einschaltkommando vorlag, zu Verfahrensschritt **1** zurückgekehrt.

[0025] Bei bereits eingeschaltetem Fernseher folgt auf Verfahrensschritt **6** der Verfahrensschritt **7**. In diesem wird überprüft, ob die Spracheingabe von dem

bereits vorher in dem Verfahrensschritt **10** vermerkten Benutzer erfolgt ist. Falls dieses der Fall ist, wird in dem Verfahrensschritt **8** das eingegebene Kommando zur Steuerung des sprachgesteuerten Systems, z. B. zu einer Menüsteuerung bzw. -navigation, verwendet. Anschließend wird, ebenso wie für den Fall, daß im Verfahrensschritt **7** eine Abweichung der Benutzer festgestellt wurde, zu Verfahrensschritt **1** zurückgekehrt.

[0026] Es sind verschiedene Abwandlungen, von diesem Ausführungsbeispiel denkbar. So kann eine Spracheingabe zum Ausschalten des Gerätes auch von einem beliebigen Benutzer akzeptiert werden. Ebenso kann ein Bedienungsbeefehl vorgesehen sein, bei dessen Eingabe durch den ersten Benutzer zukünftig auch Spracheingaben von einem zweiten Benutzer oder weiteren Benutzern akzeptiert werden.

[0027] In [Fig. 2](#) ist schematisch der Ablauf eines zweiten Ausführungsbeispiels in Form eines Flußdiagramms dargestellt. In diesem Fall erfolgt bei einem sprachgesteuerten Fernseher eine Identifizierung verschiedener Benutzer, um hieraus Benutzerprofile für diese Benutzer zu erstellen.

[0028] Die Verfahrensschritte **1** bis **5** stimmen hierbei mit dem Ausführungsbeispiel aus [Fig. 1](#) überein, wobei es in Verfahrensschritt **5** nun jedoch unbedingt erforderlich ist, den gerade sprechenden Benutzer zu identifizieren. Die verschiedenen Abzweigungen in den Verfahrensschritten **6**, **7** und **9** sind dagegen nun nicht erforderlich. Die der Spracheingabe entsprechenden Bedienbefehle werden in dem Verfahrensschritt **8** ausgeführt. Weiterhin wird in dem Verfahrensschritt **11** der identifizierte Benutzer, von dem die Eingabe stammt, zusammen mit Angaben zur aktuellen Zeit und dem Fernsehkanal oder Angaben zu der gerade laufenden Sendung abgespeichert. Diese Angaben können hierbei bereits im Fernsehgerät vorliegen oder auch mit dem Fernsehsignal als Zusatzsignal, bei analogem Fernsehsignal insbesondere in der vertikalen Austastlücke, übertragen werden. So kann beispielsweise die Uhrzeit aus einer internen Uhr verwendet werden oder aber das in Videotext übertragene Zeitsignal ausgewertet werden. Ebenso kann der Fernsehkanal unmittelbar aus dem gerade gewählten Programmplatz des Fernsehgerätes ermittelt werden oder aber aus entsprechenden Angaben im Videotext- oder VPS-Signal. Schließlich können auch Angaben zu der gerade laufenden Sendung, d.h. der Titel oder die Sparte wie z.B. Unterhaltung, Sport etc. einem bereits im Gerät vorliegenden EPG oder aber auch entsprechend übertragenen Daten entnommen werden.

[0029] Die durch die Verwendung der Sprechererkennung ermittelten Benutzerprofile können insbesondere bei der Zwischenspeicherung von TV-Pro-

grammen auf Festplatten oder ähnlichen Speichermedien, die in Fernsehern und Set-Top Boxen vorgesehen werden, eingesetzt werden. Die Genauigkeit der Analyse der Sehgewohnheiten erhöht sich hierbei deutlich durch die Erkennung des jeweiligen Benutzers. Für das Beispiel einer Familie, bei der die Kinder deutlich mehr Zeit vor dem Fernseher verbringen als die Eltern, wird daher die Festplatte nicht mehr nur mit Kinderprogrammen aufgefüllt. Vielmehr kann durch die zusätzliche Sprechererkennung die Sehgewohnheitsanalyse für mehrere Familienmitglieder getrennt erstellt werden. Der begrenzte Zwischenspeicher-Platz der Festplatte kann dann nach einem bestimmten Schlüssel auf die einzelnen Benutzer aufgeteilt werden, so daß jeder Benutzer seinen vorbestimmten Anteil an zwischengespeicherten Fernsehprogrammen erhält.

[0030] Ebenso können durch die Verwendung der Sprechererkennung ermittelte Benutzerprofile auch für die Aufzeichnung von Radioprogrammen oder andere übertragene Daten verwendet werden.

[0031] Zur Detektion der Sprachsignale kann ein einzelnes Mikrofon, aber auch ein Mikrofonarray aus zwei oder mehreren Mikrofonen vorgesehen sein. Das Mikrofonarray kann beispielsweise in einem Fernsehempfänger integriert sein. Die Mikrofone setzen die detektierten Schallsignale in elektrische Signale um, die durch Verstärker verstärkt, durch AD-Wandler in digitale Signale umgewandelt und dann einer Signalverarbeitungseinheit zugeführt werden. Diese kann den jeweiligen Aufenthaltsort des Benutzers durch eine unterschiedliche Skalierung bzw. Verarbeitung der detektierten Schallsignale berücksichtigen. Weiterhin kann auch eine Korrektur der Mikrofonensignale bezüglich der von den Lautsprechern abgegebenen Schallsignale erfolgen. Das so bearbeitete Signal wird dann dem Spracherkennung und Sprecherkennung zugeführt, wobei Algorithmen oder Hardwareeinheiten separat aber ebenso gemeinsam ausgeführt sein können. Die ermittelten Befehle sowie die Identität des Benutzers werden dann schließlich einem Systemmanager zur Steuerung des Systems zugeführt.

[0032] Die Erfindung kann bei der Sprachfernbedienung von verschiedensten Geräten der Unterhaltungselektronik, wie z. B. von TV-Geräten, Videorecordern, DVD-Spielern, Satellitenempfängern, TV-Video-Kombinationen, Audiogeräten oder kompletten Audiosystemen eingesetzt werden.

Patentansprüche

1. Verfahren zur Sprachsteuerung von einem Gerät der Unterhaltungselektronik, mit den folgenden Schritten:
Spracheingaben eines Benutzers werden in digitalisierte Sprachsignale umgewandelt;

wobei aus den digitalisierten Sprachsignalen erste Merkmale extrahiert werden, die charakteristisch für die einzelnen Laute der Sprache sind und so eine Erkennung der gesprochenen Laute ermöglichen, und wobei aus den digitalisierten Sprachsignalen zweite Merkmale extrahiert werden, die eine Charakterisierung der Stimme des jeweiligen Benutzers ermöglichen und für eine Unterscheidung der Spracheingaben von verschiedenen Benutzern verwendet werden;

dadurch gekennzeichnet, daß nach einem Sprachkommando von einem ersten Benutzer weitere Spracheingaben nur noch von diesem ersten Benutzer akzeptiert werden, indem die weiteren Spracheingaben auf charakteristische Sprachmerkmale geprüft werden und nur dann akzeptiert werden, wenn sie aufgrund dieser Merkmale demselben Sprecher zuzuordnen sind.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß ein Sprachkommando zum Einschalten des Gerätes von einem beliebigen ersten Benutzer akzeptiert wird und danach Sprachkommandos vorübergehend nur von diesem ersten Benutzer akzeptiert werden.

3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, daß ein Sprachkommando zum Ausschalten des Gerätes nur von dem ersten Benutzer akzeptiert wird und daß nach dem Ausschalten des Gerätes wieder Sprachkommandos von einem beliebigen Benutzer akzeptiert werden.

4. Verfahren nach Anspruch 2, dadurch gekennzeichnet, daß ein Sprachkommando zum Ausschalten des Gerätes von einem beliebigen Benutzer akzeptiert wird.

5. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß ein Bedienungsbefehl vorgesehen ist, nach dessen Eingabe durch den ersten Benutzer Spracheingaben von einem zweiten Benutzer akzeptiert werden.

6. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß eine Identifizierung der verschiedenen Benutzer und eine Analyse der Sehgewohnheiten der identifizierten Benutzer erfolgt, um hieraus Benutzerprofile der verschiedenen Benutzer zu erstellen.

7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß das Benutzerprofil bei einer Zwischenspeicherung von Fernsehprogrammen verwendet wird, um eine getrennte Zwischenspeicherung von bevorzugten Sendungen für verschiedene Benutzer zu ermöglichen.

8. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß das Benutzerprofil verwendet wird, um

dem identifizierten Benutzer Vorschläge für zu betrachtende Sendungen zu machen.

9. Vorrichtung mit Mittel zur Durchführung jeder der Schritte eines Verfahrens nach einem der vorhergehenden Ansprüche.

Es folgen 2 Blatt Zeichnungen

Anhängende Zeichnungen

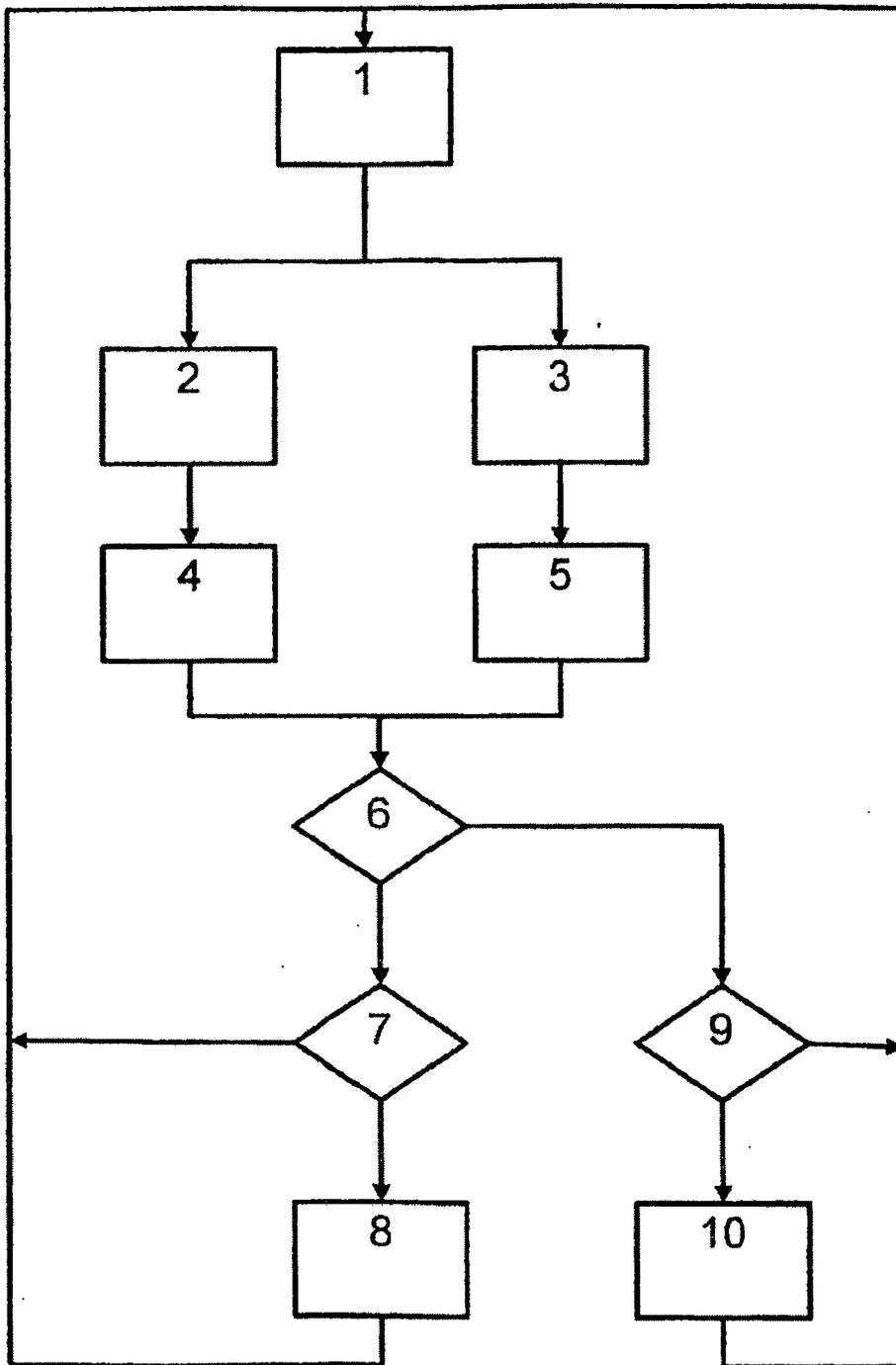


FIG. 1

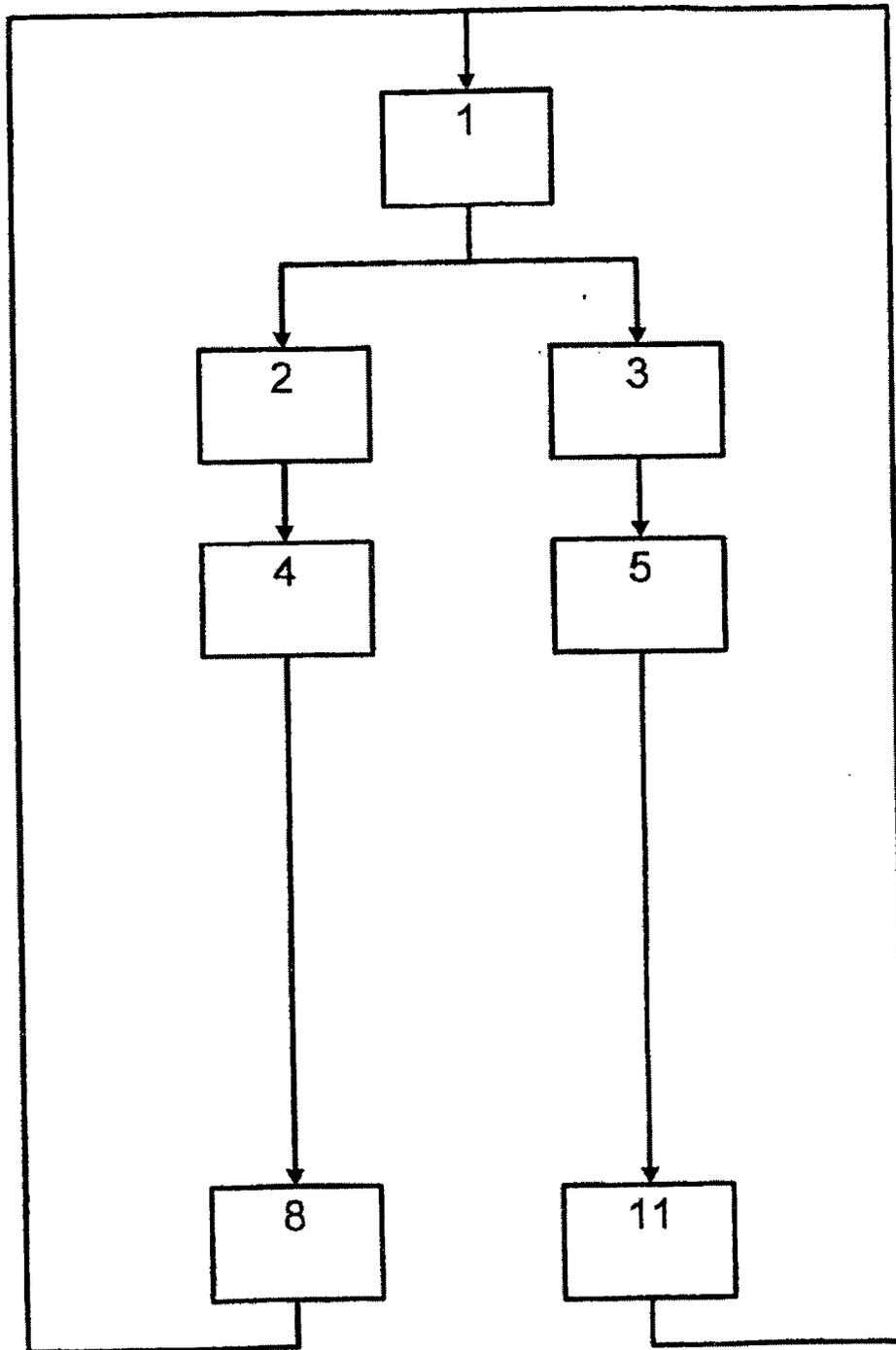


FIG. 2