



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 698 17 844 T2 2004.05.06**

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 016 078 B1**

(51) Int Cl.7: **G10L 15/22**

(21) Deutsches Aktenzeichen: **698 17 844.0**

(86) PCT-Aktenzeichen: **PCT/US98/13118**

(96) Europäisches Aktenzeichen: **98 931 545.2**

(87) PCT-Veröffentlichungs-Nr.: **WO 99/000790**

(86) PCT-Anmeldetag: **26.06.1998**

(87) Veröffentlichungstag
der PCT-Anmeldung: **07.01.1999**

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

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **03.09.2003**

(47) Veröffentlichungstag im Patentblatt: **06.05.2004**

(30) Unionspriorität:
50998 P 27.06.1997 US

(74) Vertreter:
derzeit kein Vertreter bestellt

(73) Patentinhaber:
**M.H. Segan Ltd. Partnership, Great Barrington,
Mass., US**

(84) Benannte Vertragsstaaten:
DE, GB

(72) Erfinder:
SEGAN, H., Marc, New York, US

(54) Bezeichnung: **VERFAHREN UND VORRICHTUNG ZUR SPRACHERKENNUNGSCOMPUTEREINGABE**

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

HINTERGRUND DER ERFINDUNG

1. Gebiet der Erfindung

[0001] Die vorliegende Erfindung betrifft ein Verfahren und eine Vorrichtung für eine Computer-Schnittstelle, in welcher die Stimme eines Benutzers verwendet wird, um Daten in einen Computer einzugeben. Insbesondere betrifft die vorliegende Erfindung ein Verfahren und eine Vorrichtung, in denen die Stimme eines Benutzers in Verbindung mit einer Tastatur verwendet wird, um Anweisungen oder Daten in einen Computer einzugeben. Insbesondere besonders betrifft die vorliegende Erfindung ein Verfahren und eine Vorrichtung zur Eingabe in einen Spracherkennungscomputer, bei der ein Benutzer Daten in einen Computer eingibt, indem er jedes Wort ausspricht, während er es über eine Tastatur eintippt, wobei der erste Buchstabe eines jeden gesprochenen Wortes den Computer bei der Identifizierung oder dem Erkennen der gesprochenen Worte unterstützt.

2. Beschreibung des Standes der Technik

[0002] Gegenwärtig gibt es verschiedene preisgünstige Software-Programme zur Verwendung mit Personal Computern, die eine Eingabe von Text und Anweisungen über die Stimme gestatten. Einige solcher Programme werden von Dragon Systems, IBM und Kurzweil angeboten. Diese Programme ermöglichen es, das von einem Benutzer Gesprochene zum Beispiel über eine Audio-Eingabe-Vorrichtung in einen Computer einzugeben, woraufhin die gesprochenen Worte identifiziert und für verschiedene Anwendungen, zum Beispiel für die Texteingabe oder für die Ausführung von Befehlen, verwendet werden. Wie einzusehen ist, sparen solche Mensch/Computer-Diktierprogramme Zeit des Benutzers, indem sie die Notwendigkeit, Daten über eine Tastatur für die Erzeugung von Dokumenten einzugeben, vermindern.

[0003] Zum Beispiel offenbart das Dokument US-A-5027406 ein bekanntes Spracherkennungsverfahren, das in einem Diktier-Software-Programm verwendet wird.

[0004] Unter Bezugnahme auf **Fig. 1** ist ein Diagramm der Schritte, die in einem Programm zur Ausführung einer Spracheingabe in einen Computer nach dem Stand der Technik verwendet werden, abgebildet. Wie gezeigt, ist das Programm gemäß dem Stand der Technik in fünf getrennte Schritte unterteilt. Der erste Schritt schließt ein, dass der Benutzer in der Nähe des Computers spricht, so dass ein Klangdetektor das gesprochene Audio-Signal empfangen kann. Der Computer bestimmt dann, wann jedes Wort beginnt. Für jedes gesprochene Wort wird eine Suche in einem Wörterbuch ausgeführt, um mögliche Übereinstimmungen zwischen den gespeicherten

Wörterbuchwörtern und jedem gesprochenen Wort zu lokalisieren. Aus den potenziellen Übereinstimmungen wird dann die beste Übereinstimmung mit dem Wörterbuchwort ausgewählt, und das ausgewählte Wort wird an eine Anwendung weitergeleitet. [0005] Unglücklicherweise besitzen solche Programme nach dem Stand der Technik einige Nachteile. Besonders die Fehlerrate solcher Programme bleibt hoch. Darüber hinaus bestehen Probleme bei der kontinuierlichen Spracheingabe, so dass ein unzweckmäßiger Spracheingabestil erforderlich ist, indem an Intervallen pausiert werden muss, um es dem Programm zu gestatten, die gesprochenen Wörter zu identifizieren und die Informationen zu verarbeiten. Auch müssen solche Programme nach dem Stand der Technik vor der Verwendung trainiert werden. Anders ausgedrückt muss ein Benutzer vor der Verwendung Zeit zur Aufnahme eines umfangreichen Muster-Vokabulars für die Maschine aufwenden, damit diese sie mit den Eingaben vergleichen kann.

[0006] Als ein Ergebnis der Schwierigkeiten von Programmen nach dem Stand der Technik stehen die Benutzer der Situation gegenüber, dass sie sich für eine genaue und bequeme Texteingabe weiterhin auf alte Verfahren verlassen müssen, welche entweder beinhalten, dass jemand ein Diktat aufnehmen und dann die gesprochenen Wörter über eine Tastatur in den Computer eingeben muss, oder schnell Maschineschreiben kann. Ausserdem führt die fortgesetzte und regelmäßige Verwendung einer Tastatur zu einem Risiko von Verletzungen und Ermüdung durch wiederholte Belastung.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0007] In den Zeichnungen ist:

[0008] **Fig. 1** ein Blockdiagramm, welches die Schritte eines Stimmeingabeverfahrens nach dem Stand der Technik darstellt;

[0009] **Fig. 2** ein Blockdiagramm, welches die Schritte einer Ausführungsform des Verfahrens und der Vorrichtung zur Eingabe in einen Spracherkennungs-Computer gemäß der vorliegenden Erfindung darstellt; und

[0010] **Fig. 3** ein Blockdiagramm, welches die Schritte einer weiteren Ausführungsform des Verfahrens und der Vorrichtung zur Eingabe in einen Spracherkennungs-Computer gemäß der vorliegenden Erfindung darstellt.

DETAILLIERTE BESCHREIBUNG DER GEGENWÄRTIG BEVORZUGTEN AUSFÜHRUNGSFORMEN

[0011] Die vorliegende Erfindung verwendet eine bekannte Stimmerkennungstechnologie, um es dem Benutzer zu gestatten, den Anfangsbuchstaben eines jeden gesprochenen Wortes über eine Tastatur, im Wesentlichen gleichzeitig mit dem gesprochenen Wort, einzugeben, wodurch die Verarbeitungsleis-

tung und die Zeit, die von einem Computer benötigt werden, verringert werden. Die Erfindung überwindet die Nachteile des Standes der Technik in mehreren Beziehungen. Erstens wird ein kontinuierliches Sprechen möglich, da dem Computer – durch Eingabe des Anfangsbuchstabens eines jeden Wortes mittels Tastatur – exakt mitgeteilt, wann ein neues Wort beginnt, wodurch die Verarbeitungsleistung, die in Systemen nach dem Stand der Technik erforderlich war, um zu bestimmen, wann ein Wort geendet und das nächste begonnen hat, in großem Ausmaß verringert wird. Des Weiteren wird die mögliche Auswahl von Wörtern in großem Ausmaß verringert, da der Anfangsbuchstabe über die Eingabe mittels Tastatur bekannt ist. Diese Eigenschaft verringert des Weiteren die erforderliche Verarbeitungsleistung und -zeit, da weniger Wörter aus dem Wörterbuch lokalisiert und abgefragt werden müssen. Diese Eigenschaft erhöht auch die Genauigkeit des ausgewählten oder abgefragten Wortes, da der Anfangsbuchstabe des gesprochenen Wortes an den Computer geliefert wird.

[0012] Zusätzlich zu den vorangegangenen Vorteilen benötigt die vorliegende Erfindung deutlich weniger Eintippen als die standardmäßige Eingabe über eine Tastatur. Als ein Ergebnis können kleinere Tastaturen im Vergleich zu Tastaturen in Standard-Größen verwendet werden, wodurch die Oberfläche des Benutrer-Arbeitsplatzes vergrößert wird, der früher von einer Tastatur in Standard-Größe beansprucht wurde. Überdies ist die vorliegende Erfindung besonders gut geeignet für die Verwendung von Computern im Palm-Format oder Persönliche digitale Assistenten (Personal Digital Assistants = PDA's). Außerdem wird, da weniger Tippen für die Dateneingabe erforderlich ist, das Risiko verringert, dass Verletzungen, Müdigkeit oder Fehler auftreten.

[0013] Als ein Ergebnis der vorliegenden Erfindung können sprecherunabhängige Programme leichter hergestellt werden, da die Notwendigkeit, sich auf die Fähigkeit des Computers zu verlassen, Abweichungen in der Aussprache zu unterscheiden, beträchtlich vermindert wird.

[0014] Die vorliegende Erfindung ist in **Fig. 2** dargestellt. Wie gezeigt, werden die Schritte **1A** und **1B** im Wesentlichen gleichzeitig miteinander ausgeführt. In der Alternative kann ein Puffer-Register eingesetzt werden, um temporär entweder das gesprochene Wort oder den getippten Buchstaben zu speichern, während die Eingabe des jeweils anderen von gesprochenem Wort oder eingetipptem Buchstaben erwartet wird. Wenn sowohl ein gesprochenes Wort als auch ein Buchstabe in den Computer eingegeben worden sind, wird eine Suche in den Wörterbuch-Einträgen ausgeführt, beginnend mit dem getippten Buchstaben, um mögliche Übereinstimmungen zu lokalisieren. Die beste Übereinstimmung wird dann ausgewählt, und das ausgewählte Wort wird an eine Anwendung weitergeleitet.

[0015] Zusätzlich zu dem oben Erwähnten zeigt **Fig. 3**, dass die vorliegende Erfindung einzelne Wör-

ter oder ein Vielzahl von Wörtern spezifischen Funktionen zuordnen oder für diese bezeichnen kann, welche ausgeführt werden, wenn das bestimmte Wort gesprochen wird, ohne dass der Anfangsbuchstabe des Wortes getippt wird, wie in den Schritten **5** und **6** von **Fig. 3**. Als eine Veranschaulichung kann dem Wort "löschen" die Funktion zugewiesen werden, das zuletzt getippte Wort zu löschen, wenn das Wort "löschen" gesprochen wird, ohne den Buchstaben "1" zu tippen. Wenn daher der Computer ein Wort auf Basis des gesprochenen Wortes und des getippten Anfangsbuchstaben des gesprochenen Wortes nicht korrekt identifiziert, kann der Benutzer das nicht korrekte Wort anschlagen oder löschen in dem er das Wort "löschen" ausspricht. Selbstverständlich wird, wenn das Wort "löschen" eingegeben werden soll, das Wort ausgesprochen und der Buchstabe "1" auf der Tastatur gedrückt. Eine solche Funktion kann an jedem beliebigen nicht korrekten Wort ausgeführt werden, indem einfach der Cursor zu dem nicht korrekten Wort bewegt und "löschen" gesagt wird. Schritt **5** umfasst den Schritt, dass das gesprochene Audiosignal mit einer Liste vorbestimmter Funktionen verglichen wird. Optional kann Schritt **5** die Schritte umfassen, dass bestimmt wird, ob der Benutzer einen Anfangsbuchstaben tippt, und dass dann das gesprochene Audiosignal mit der Liste vorbestimmter Funktionen verglichen wird.

[0016] Die Erfindung wird durch die oben beschriebenen Ausführungsformen, welche nur als Beispiele gezeigt werden, nicht eingeschränkt, sondern kann auf verschiedene Weise abgewandelt werden, wie für den Fachmann leicht ersichtlich ist. Zum Beispiel können, anstatt dass nur der Anfangsbuchstabe jedes gesprochenen Wortes getippt wird, der erste und der letzte Buchstabe getippt werden, oder alternativ, die ersten beiden Buchstaben getippt werden, wodurch die Zeit und Energie für die Computerverarbeitung weiter verringert wird.

Patentansprüche

1. Verfahren zur Erleichterung der Identifizierung von gesprochenen Wörtern durch eine Spracherkennungs-Computeranwendung, die auf eine Speichervorrichtung zugreift, welche Wörterbucheinträge von gespeicherten Wörterbuch-Wörtern enthält und welche identifizierte Wörter in eine gleichzeitig laufende Computeranwendung eingibt, wobei das Verfahren die Schritte umfasst, dass:

- a. ein Audiosignal empfangen wird, das ein von einem Benutzer gesprochenes Wort darstellt;
- b. eine von einem Benutzer über Tasten eines von einem Benutzer bedienbaren Tastenfeldes eingegebene Eingabe empfangen wird, welche zumindest den ersten Buchstaben des gesprochenen Wortes darstellt;
- c. nach Schritt a. und b. automatisch eine Übereinstimmung zwischen dem gesprochenen Wort und den Wörterbucheinträgen durch Vergleichen des Au-

diosignals nur mit jenen Wörterbucheinträgen ausgewählt wird, die den zumindest ersten Buchstaben enthalten; und

d. danach die Übereinstimmung in die gegenwärtig laufende Computer-Anwendung eingefügt wird, ohne weitere Benutzereingaben zu erfordern.

2. Verfahren nach Anspruch 1, worin die Schritte a. und b. im Wesentlichen zeitgleich ausgeführt werden.

3. Verfahren nach Anspruch 1, worin einer der Schritte a. und b. vor dem anderen der Schritte a. und b. ausgeführt wird, und das Computer-Schnittstellen-Verfahren des Weiteren den Schritt umfasst, dass der eine der Schritte a. und b. in einem Puffer-Register gespeichert wird, bis der andere der Schritte a. und b. ausgeführt wird.

4. Verfahren nach einem der Ansprüche 1–3, des Weiteren umfassend die Schritte, dass:

e. bestimmt wird, ob das gesprochene Audiosignal eine vorbestimmte Funktion umfasst;

f. die vorbestimmte Funktion ausgeführt wird, wenn in Schritt e. bestimmt wird, dass das gesprochene Audiosignal die vorbestimmte Funktion umfasst; und

g. Schritt c. ausgeführt wird, wenn die gesprochene Audio-Schnittstelle nicht die vorbestimmte Funktion ist.

Es folgen 3 Blatt Zeichnungen

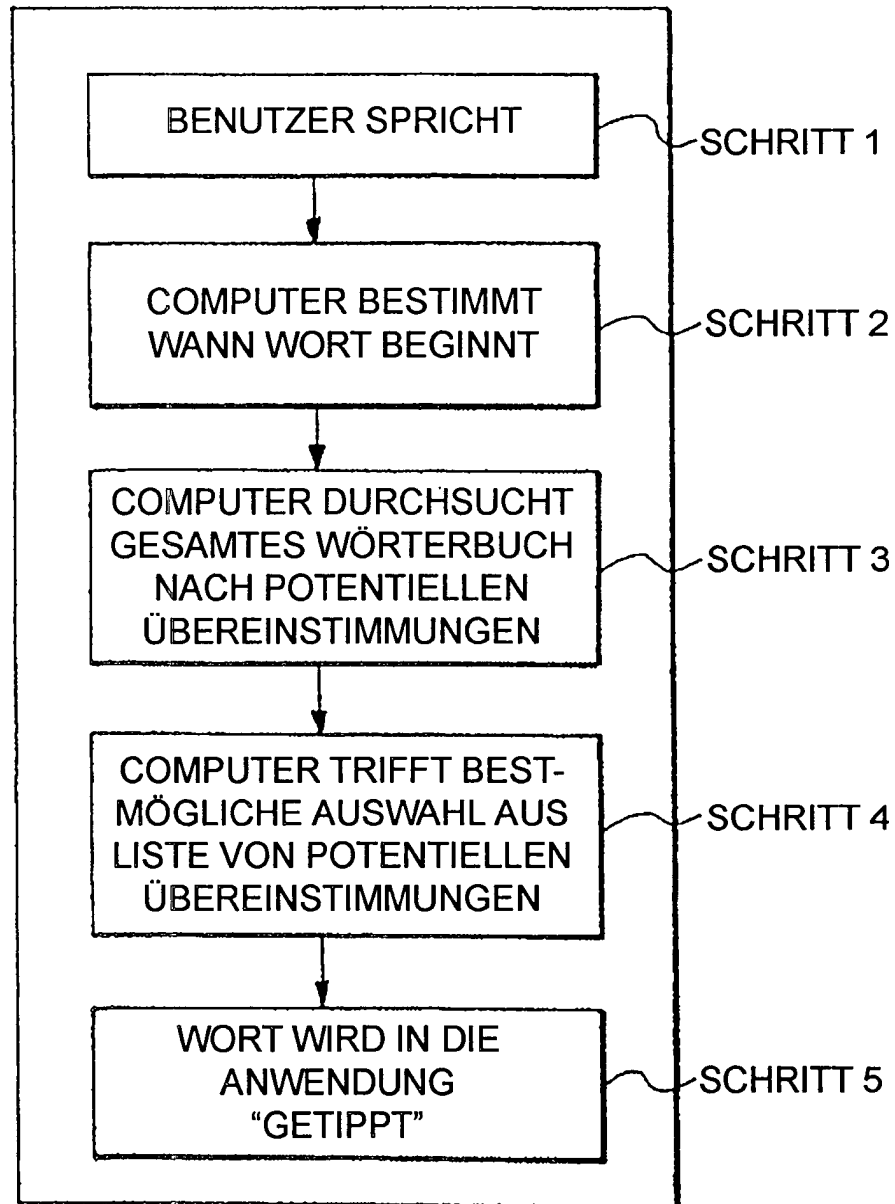


FIG. 1

STAND DER TECHNIK

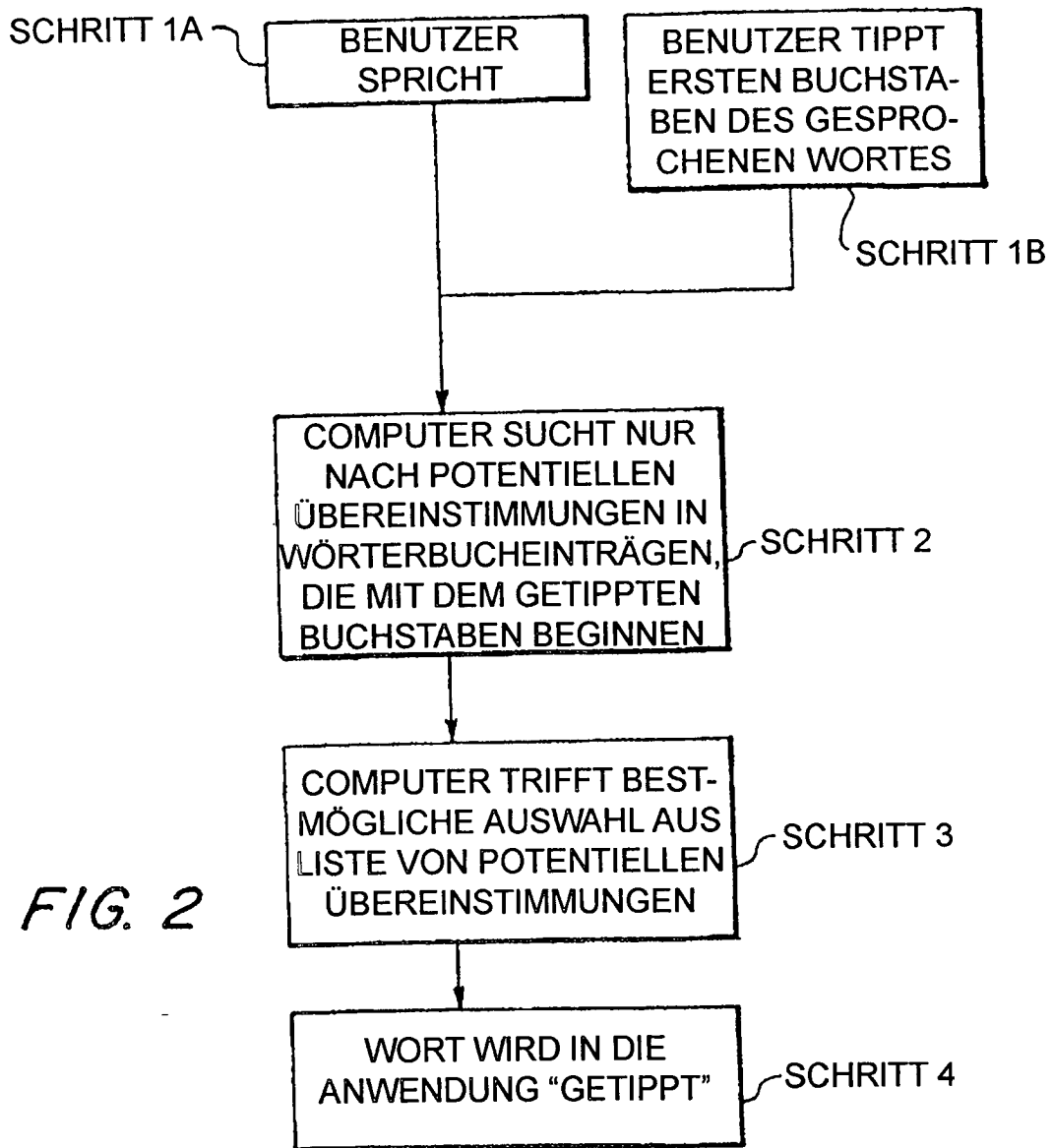


FIG. 2

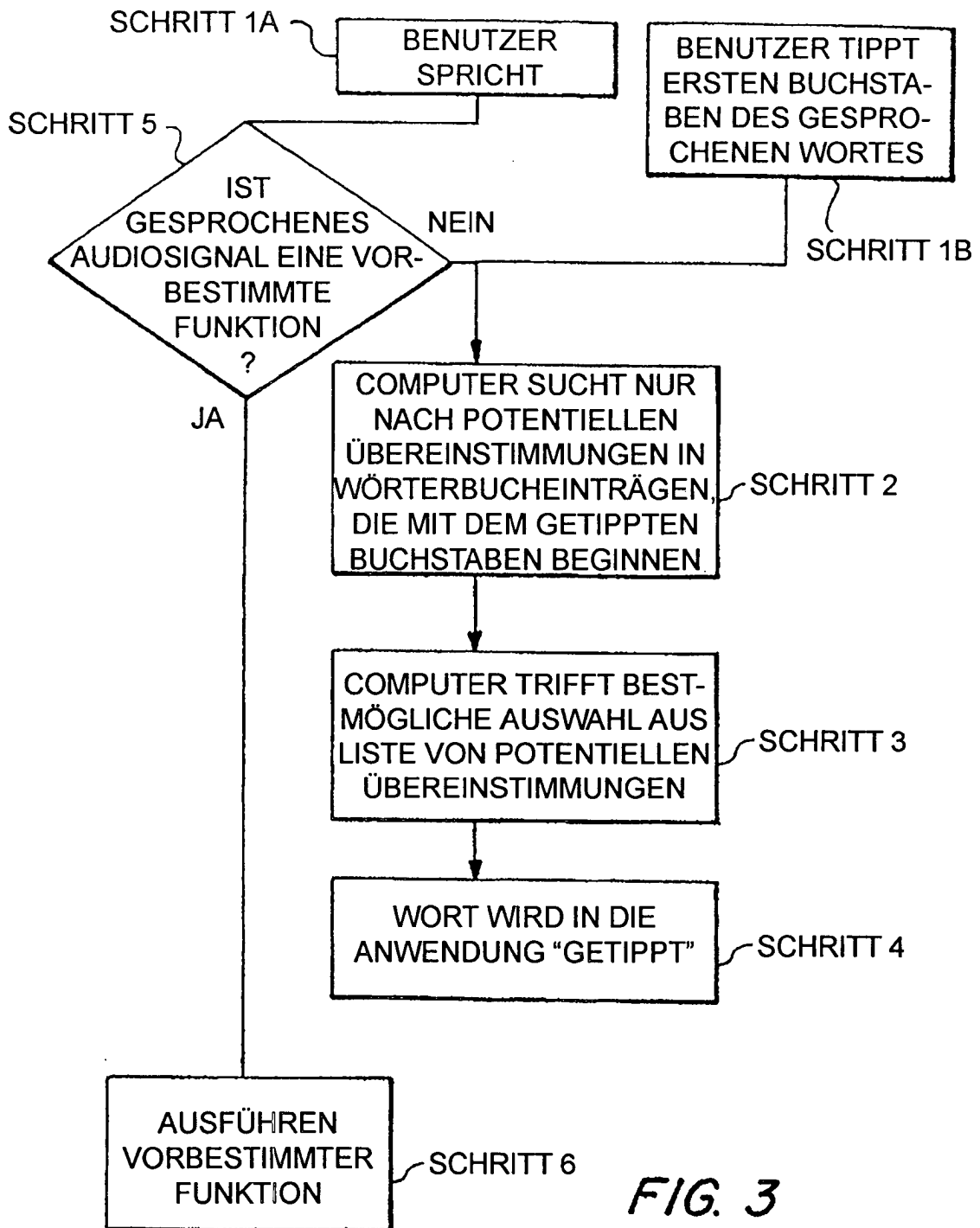


FIG. 3