

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
21. Dezember 2000 (21.12.2000)

PCT

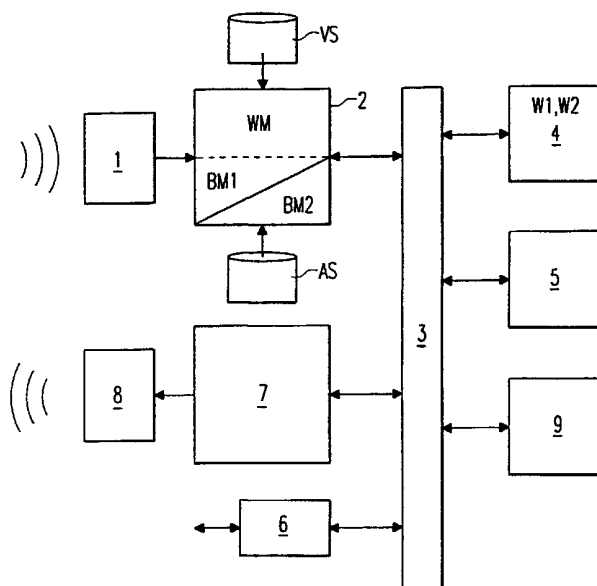
(10) Internationale Veröffentlichungsnummer
WO 00/77771 A2

- (51) Internationale Patentklassifikation⁷: **G10L** (72) Erfinder; und
(75) Erfinder/Anmelder (nur für US): **SCHNEIDER, Peter**
(21) Internationales Aktenzeichen: **PCT/DE00/01734** [DE/DE]; Enhuberstr. 12, D-80333 München (DE).
(22) Internationales Anmeldedatum: **29. Mai 2000 (29.05.2000)** (74) Gemeinsamer Vertreter: **INFINEON TECHNOLOGIES AG**; Lambsdorff & Lange, Dingolfinger Strasse 6, 81673 München (DE).
(25) Einreichungssprache: **Deutsch** (81) Bestimmungsstaaten (national): **CN, JP, KR, US.**
(26) Veröffentlichungssprache: **Deutsch** (84) Bestimmungsstaaten (regional): **europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).**
(30) Angaben zur Priorität: **199 26 468.6** **10. Juni 1999 (10.06.1999)** **DE**
(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): **INFINEON TECHNOLOGIES AG [DE/DE]; St.-Martin-Str. 53, D-81541 München (DE).** Veröffentlicht:
— Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

[Fortsetzung auf der nächsten Seite]

(54) Title: **VOICE RECOGNITION METHOD AND DEVICE**

(54) Bezeichnung: **SPRACHERKENNUNGSVERFAHREN UND -EINRICHTUNG**



(57) Abstract: The invention relates to a method for interactive voice recognition according to which a word uttered by a user is analyzed. If said word cannot be unequivocally assigned to a word in a defined vocabulary, several possible words contained in said vocabulary are assigned to the word. To each word forming part of this preselection a differentiation letter is assigned by means of which said word can be distinguished from the others. The user is then requested to pronounce the letter assigned to the word the device is searching for and said word is then determined by voice recognition of the differentiation letter.

[Fortsetzung auf der nächsten Seite]



WO 00/77771 A2



Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(57) Zusammenfassung: Bei einem Verfahren zur interaktiven Spracherkennung wird ein von einem Benutzer gesprochenes Wort analysiert. Sofern keine eindeutige Zuordnung zu einem Wort eines vorgegebenen Vokabulars möglich ist, werden dem Wort mehrere mögliche Wörter des Vokabulars zugeordnet. Zu jedem Wort dieser Vorauswahl wird ein Entscheidungsbuchstaben bestimmt, der ein Wort von den anderen Wörtern der Vorauswahl unterscheidbar macht. Der Benutzer wird aufgefordert, den Entscheidungsbuchstaben zu dem gesuchten Wort auszusprechen. Durch Spracherkennung des Entscheidungsbuchstabens wird das gesuchte Wort ermittelt.

Beschreibung

Spracherkennungsverfahren und -einrichtung

- 5 Die Erfindung betrifft ein Verfahren und eine Einrichtung zur Spracherkennung eines Wortes.

In vielen Bereichen der Technik finden Spracherkennungssysteme zunehmende Verbreitung.

10

Bei Diktiersystemen werden Spracherkennungstechniken für die automatische Erstellung von geschriebenem Text eingesetzt. Derartige Diktiersysteme beruhen auf der Erkennung einzelner Wörter oder Silben. Neben der Wort- oder Silbenerkennung
15 weisen sie häufig einen Buchstabiermodus auf, der den Benutzer bei Nichterkennung eines Wortes dazu auffordert, das Wort Buchstabe für Buchstabe auszusprechen.

20

Andere bekannte Spracherkennungsanwendungen basieren a-priori auf einer buchstabenweisen Eingabe eines Wortes. Derartige Systeme werden als Buchstabiererkenner bezeichnet. Buchstabiererkenner werden beispielsweise in Navigationssystemen für Kraftfahrzeuge mit Spracheingabe von Zielinformation eingesetzt. Das Navigationssystem muß eine sehr große Anzahl von
25 teilweise ähnlich klingenden Wörtern (Städtenamen, Straßennamen, Hotel- und Gaststättennamen, usw.) unterscheiden können, was durch eine buchstabenweise Eingabe des Wortes mit hinreichender Sicherheit gewährleistet werden kann. Nachteilig ist jedoch, daß für das Buchstabieren eine relativ hohe Konzentration erforderlich ist, die beim Bewegen eines Kraftfahrzeugs nicht immer aufgebracht werden kann.
30

35

Der Erfindung liegt die Aufgabe zugrunde, ein Verfahren und eine Einrichtung zur Spracherkennung eines Wortes anzugeben, das bzw. die im Falle einer nicht eindeutigen Erkennung eines Wortes in einem Wort-Erkennungsmodus einen für einen Benutzer

bequemen Ablauf zum Auffinden des gesuchten Wortes durchführt.

5 Zur Lösung der Aufgabe sind die Merkmale der Ansprüche 1 und 8 vorgesehen.

Die Erfindung beruht auf der Erkenntnis, daß es in der Regel nicht erforderlich ist, den Benutzer das im Wort-Erkennungsmodus nicht eindeutig erkannte Wort im Buchstabier-Erkennungsmodus von Anfang an buchstabieren zu lassen. Vielmehr
10 reicht es aus, die begrenzte Anzahl der in der Vorauswahl enthaltenen Wörter anhand von geeignet gewählten Entscheidungsbuchstaben unterscheidbar zu machen und dann das gesuchte Wort durch gezieltes Abfragen des dem gesuchten Wort zugeordneten Entscheidungsbuchstabens im Buchstabier-Erkennungsmodus zu ermitteln.
15

Vorzugsweise werden bei der Abfrage des Entscheidungsbuchstabens die zuvor bezüglich der Wörter der Vorauswahl bestimmten
20 Entscheidungsbuchstaben dem Benutzer von dem Spracherkennungssystem vorgeschlagen, wodurch die Interaktivität des Systems erhöht wird.

Obgleich die Abfrage grundsätzlich beispielsweise auch optisch erfolgen kann, ist es bei vielen Anwendungen zweckmäßig, eine akustische Abfrage vorzusehen.
25

In der Praxis kann es vorkommen, daß aufgrund von plötzlich auftretenden Umgebungsgeräuschen oder einer zunächst undeutlichen Aussprache des Benutzers eine Wiederholung des gesprochenen Wortes zweckmäßig erscheint. Eine vorteilhafte Ausgestaltung des erfindungsgemäßen Verfahrens kennzeichnet sich daher dadurch, daß der Benutzer zur Wiederholung des gesprochenen Wortes aufgefordert wird, sofern die Anzahl der in der
30 Vorauswahl enthaltenen Wörter einen vorbestimmten Grenzwert überschreitet. Bei der Wiederholung des Wortes können weniger Störungen durch Umgebungsgeräusche vorliegen und der Be-

nutzer bemüht sich erfahrungsgemäß auch um eine deutlichere Aussprache, so daß im Ergebnis zumeist eine für die folgenden Abfrage- und Buchstabiererkennungsschritte günstigere Vorauswahl als beim ersten Versuch zur Verfügung steht.

5

Bei dem Buchstabier-Erkennungsmodus kann es sich um einen Alphabetwort-Erkennungsmodus oder einen Buchstaben-Erkennungsmodus handeln. Im ersten Fall kann der Bedienungskomfort in vorteilhafter Weise dadurch erhöht werden, daß mehrere unterschiedliche Alphabetwörter (z.B. "Anton", "Alpha", "Alfred" für den Buchstaben a) einem einzelnen Buchstaben zugeordnet sind. Der Benutzer hat dann mehrere Möglichkeiten, einen Buchstaben im Alphabetwort-Erkennungsmodus zu benennen.

10

15 Mit besonderem Vorteil kommt die erfindungsgemäße Einrichtung in einem Navigationssystem für ein Fortbewegungsmittel, insbesondere Kraftfahrzeug (Kfz) zum Einsatz, da aufgrund der einfachen Eingabeprozedur von Zielinformation in das System die Aufmerksamkeit des Fahrers nur in sehr geringem Maße vom Straßenverkehr abgelenkt wird.

20

Weitere vorteilhafte Ausgestaltungen der Erfindung sind in den Unteransprüchen angegeben.

25 Die Erfindung wird im folgenden anhand eines Ausführungsbeispiels unter Bezugnahme auf die Zeichnung erläutert; in dieser zeigt:

30

Fig. 1 ein Blockschaltbild einer erfindungsgemäßen Einrichtung;

Fig. 2 ein Blockdiagramm zur beispielhaften Erläuterung des Ablaufs eines erfindungsgemäßen Verfahrens; und

35 Fig. 3 eine tabellarische Darstellung zur Erläuterung des Wort-Erkennungsmodus und der beiden Buchstabier-Erkennungsmoden.

Fig. 1 zeigt in Form eines Beispiels ein Blockschaltbild einer Einrichtung zur Durchführung des erfindungsgemäßen Verfahrens. Einer Mikrophon/Verstärkeranordnung 1 ist ein
5 Spracherkenner 2 nachgeschaltet, der sowohl in einem Wort-Erkennungsmodus (WM) als auch in einem Buchstabier-Erkennungsmodus (BM1 oder BM2) betrieben werden kann. Ein Ausgang des Spracherkenners 2 steht mit einem Daten- und Steuerbus 3 in bidirektionaler Verbindung. An den Daten- und Steuerbus 3
10 sind ferner ein Wortspeicher 4, ein Analysator/Entscheider 5, eine Schnittstelle 6 und ein Sprachgenerator 7 angeschlossen. Ein Ausgang des Sprachgenerators 7 steht mit einer Verstärker/Lautsprecheranordnung 8 in Verbindung. Ein Microcontroller 9 ist ebenfalls mit dem Daten- und Steuerbus 3 verbunden
15 und dient zur Ablaufsteuerung des gesamten Systems.

Der Spracherkenner 2 hat Zugriff auf einen Vokabularspeicher VS, in dem die zu erkennenden Wörter sowie deren zugeordnete Klangbilder in Form von Referenz-Sprachsignalen abgelegt
20 sind. Ferner greift der Spracherkenner 2 auf einen Alphabetspeicher AS zu, in dem die Buchstaben des Alphabets sowie deren zugeordnete Klangbilder (in Form eines Wortalphabets und/oder in Form von gesprochenen Buchstaben) ebenfalls als Referenz-Sprachsignale abgelegt sind. In beiden Fällen können
25 die Klangbilder vom Benutzer trainierbar sein, wodurch eine benutzerangepaßte Spracherkennung realisiert wird. Der Vokabularspeicher VS kann z.B. eine wechselbare CD-ROM sein. Beide Speicher VS, AS können auch über den Daten- und Steuerbus 3 mit dem Spracherkenner 2 in Verbindung stehen.

30

Der Ablauf der erfindungsgemäßen interaktiven Spracherkennung wird in beispielhafter Weise anhand der Fig. 2 und 3 erläutert.

35 Zunächst befindet sich der Spracherkenner 2 im Wort-Erkennungsmodus WM. Der Benutzer spricht ein zu erkennendes Wort, beispielsweise "Padua" für die italienische Stadt Padua.

Basierend auf dem gesprochenen Wort "Padua" wird mittels der Mikrophon/Verstärkeranordnung 1 ein elektrisches Sprachsignal S("Padua") erzeugt.

- 5 Das elektrische Sprachsignal S("Padua") wird nachfolgend im Spracherkenner 2 zur Erkennung des dem Sprachsignal zugrundeliegenden Wortes analysiert. Zur Analyse des elektrischen Sprachsignals S("Padua") kann ein Korrelator verwendet werden, der das elektrische Sprachsignal S("Padua") mit dem im
- 10 Vokabularspeicher VS abgespeicherte Referenz-Sprachsignalen (Klangbildern) vergleicht.

Als erkanntes Wort wird dasjenige Wort W1 aus dem Vokabularspeicher VS bestimmt, dessen Referenz-Sprachsignal die größte

15 Korrelation mit dem erhaltenen, elektrischen Sprachsignal S("Padua") aufweist. Ferner wird ermittelt, mit welcher Sicherheit bzw. Wahrscheinlichkeit die Erkennung korrekt ist. Letzteres kann beispielsweise durch die Berechnung einer Zuverlässigkeitsinformation oder Vertrauenszahl $P(W1)$ erfolgen,

20 welche ein Maß für die Richtigkeit der Erkennung (d.h. $Padua = W1$) darstellt.

Die Vertrauenszahl $P(W1)$ kann in vielfältiger Weise erzeugt werden. Sie kann beispielsweise in Abhängigkeit von dem Absolutwert der Korrelation des elektrischen Sprachsignals

25 S("Padua") mit dem Referenz-Sprachsignal des erkannten Wortes W1 und/oder von einem Abstand zwischen dem maximalen Korrelationswert (bzgl. des erkannten Wortes W1) und dem nächstkleineren Korrelationswert (bzgl. des nächstähnlich klingenden

30 Wortes W2 aus dem Vokabular) gebildet werden. Während ersteres ein Maß für die absolute "Klangübereinstimmung" darstellt, ist letzteres ein Maß für die "Verwechslungsgefahr" zwischen W1 und W2.

- 35 Bei hoher absoluter "Klangübereinstimmung" und geringer "Verwechslungsgefahr" (es können auch noch weitere Parameter hinzugezogen werden) kann mit hoher Wahrscheinlichkeit von einer

korrekten Worterkennung ausgegangen werden. In diesem Fall nimmt das System an, daß das gesprochene Wort richtig erkannt wurde und gibt das erkannte Wort W1 als Ergebnis aus. Dem Benutzer wird auf akustischem Wege, z.B. über den Sprachgenerator 7 und die Verstärker/Lautsprecheranordnung 8 oder auf optischem Wege über ein Display die Möglichkeit gegeben, das Erkennungsergebnis (W1) zu kontrollieren und gegebenenfalls dem System als fehlerhafte Erkennung anzuzeigen.

10 Nicht selten wird jedoch der Fall auftreten, daß zwei oder mehrere Worte W1, W2, ... des Vokabulars gemäß der vorstehend beschriebenen Analyse als Erkennungsergebnis in Frage kommen, d.h. nicht mit hinreichender Sicherheit eine eindeutige Zuordnung im Korrelationsschritt erreicht wird.

15 In diesem Fall werden die gefundenen Worte W1, W2, ... unter Steuerung des Mikrocontrollers 9 in dem Wortspeicher 4 als Vorauswahl abgespeichert. In dem Wortspeicher 4 können auch die zugehörigen Zuverlässigkeitszahlen $P(W1)$, $P(W2)$, ... gespeichert werden.

20 In einem nächsten Schritt wird zu jedem gefundenen Wort W1, W2, ... ein in dem entsprechenden Wort enthaltener Entscheidungsbuchstabe bestimmt. Die Bestimmung der Entscheidungsbuchstaben erfolgt durch einen Vergleich der Wörter W1, W2, ... im Analysator/Entscheider 5.

Beispielsweise sei W1 das gesuchte Wort Padua und W2 das ähnlich klingende Wort Pavia. Die beiden Worte W1, W2 sind in den ersten beiden Buchstaben sowie im letzten Buchstaben identisch, können jedoch anhand ihrer dritten oder vierten Buchstaben auseinandergehalten werden. Mithin wird Padua der Entscheidungsbuchstabe d (oder auch u) zugeordnet und Pavia wird der Entscheidungsbuchstabe v (oder auch i) zugeordnet.

30 Sofern - wie im vorliegenden Beispiel - mehrere Möglichkeiten für die Bestimmung eines Entscheidungsbuchstabens bestehen, kann vorgesehen sein, daß Entscheidungsbuchstaben ausgewählt

werden, die einen möglichst großen klanglichen Abstand zueinander haben.

5 Nachfolgend wird der Spracherkenner 2 in einen Buchstabiermodus (BM1 oder BM2) umgeschaltet und es erfolgt eine Abfrage der Entscheidungsbuchstaben zu jedem gefundenen Wort W1, W2.

Die Abfrage kann auf akustischem Wege über den Sprachgenerator 7 und die Verstärker/Lautsprecheranordnung 8 erfolgen.
10 Im Buchstaben-Erkennungsmodus BM1 kann dem Benutzer das Erkennungsproblem beispielsweise über eine Sprachausgabe beschrieben werden, indem ihm als Alternative angeboten wird: "Sagen Sie "de" für Padua oder "vau" für Pavia". Entsprechend kann im Wortalphabet-Erkennungsmodus BM2 die Aufforderung
15 lauten: "Sagen Sie "Dora" für Padua oder "Viktor" für Pavia".

Die Aufforderung zur Nennung des richtigen Entscheidungsbuchstabens kann auch auf optischem Wege stattfinden, beispielsweise indem über ein Display eine Liste der gefundenen Wörter
20 W1, W2, ... angezeigt wird und die von dem Analysator/Entscheider 5 bestimmten Entscheidungsbuchstaben durch Aufblenden oder andere optische Maßnahmen für den Benutzer erkennbar gemacht werden.

25 Im nächsten Schritt beantwortet der Benutzer die Anfrage durch Aussprechen des dem richtigen Wort W1 zugeordneten Entscheidungsbuchstabens, also "de" (im Buchstaben-Erkennungsmodus BM1) oder "Dora" (im Alphabetwort-Erkennungsmodus BM2).

30 Der gesprochene Buchstaben wird mittels der Mikrophon/Verstärkeranordnung 1 in ein elektrisches Sprachsignal S("de") bzw. S("Dora") umgesetzt. Dieses elektrische Sprachsignal wird nachfolgend im Spracherkenner 2 zur Erkennung des dem Sprachsignal zugrundeliegenden Buchstabens d mittels eines
35 Korrelators analysiert. Dabei werden im Alphabetspeicher AS abgelegte Referenz-Sprachsignale (Klangbilder) der entspre-

chenden gesprochenen Buchstaben oder Alphabetwörter mit dem elektrischen Sprachsignal S("de") bzw. S("Dora") verglichen.

In einem optionalen, abschließenden Schritt kann vorgesehen
5 sein, daß das daraufhin identifizierte Wort W1 = Padua dem Benutzer zur Bestätigung über den Sprachgenerator 7 und die Verstärker/Lautsprecheranordnung 8 vorgesprochen oder auf dem Display angezeigt wird. Bei einer Bestätigung durch den Be-
nutzer wird das Wort W1 dem System (z.B. Kfz-Navigations-
10 system oder Computer) über die Schnittstelle 6 mitgeteilt, so daß das System daraufhin von dem erkannten Wort abhängige Steuermaßnahmen einleiten kann.

Patentansprüche

1. Verfahren zur interaktiven Spracherkennung eines Wortes durch ein Spracherkennungssystem, mit den Schritten:

- 5 - in einem Wort-Erkennungsmodus (WM)
-- Umsetzen eines gesprochenen Wortes ("Padua") in ein elektrisches Wort-Sprachsignal (S("Padua")) und
-- Analysieren des Wort-Sprachsignals (S("Padua")) zur Erkennung des Wortes (Padua) aus einem Vokabular (VS) vorgegebener Wörter;
10 - sofern eine eindeutige Zuordnung des Wort-Sprachsignals (S("Padua")) zu einem Wort des Vokabulars (VS) nicht getroffen werden kann, Zusammenstellen einer Vorauswahl von mehreren Wörtern (W1, W2) des Vokabulars, unter denen sich
15 das gesprochenen Wort ("Padua") mit hinreichender Sicherheit befindet;
- Bestimmen mindestens eines Entscheidungsbuchstabens zu jedem Wort (W1, W2) der Vorauswahl, welcher ein Wort von den anderen Wörtern der Vorauswahl unterscheidbar macht;
20 - in einem Buchstabier-Erkennungsmodus (BM1, BM2)
-- Abfragen der Entscheidungsbuchstaben zu den Wörtern der Vorauswahl;
-- Umsetzen einer Buchstabier-Spracheingabe ("de", "Dora") in ein zugeordnetes elektrisches Buchstabier-Sprachsignal
25 (S("de"), S("Dora")), und
-- Analysieren des Buchstabier-Sprachsignals (S("de"), S("Dora")) zur Erkennung des Entscheidungsbuchstabens (d).

2. Verfahren nach Anspruch 1,

- 30 d a d u r c h g e k e n n z e i c h n e t,
daß bei der Abfrage die zuvor bestimmten Entscheidungsbuchstaben dem Benutzer von dem Spracherkennungssystem vorgeschlagen werden.

35 3. Verfahren nach Anspruch 1 oder 2,

- d a d u r c h g e k e n n z e i c h n e t,
daß die Abfrage akustisch erfolgt.

4. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
daß der Benutzer zur Wiederholung des gesprochenen Wortes
5 ("Padua") aufgefordert wird, sofern die Anzahl der in der
Vorauswahl enthaltenen Wörter (W1, W2) einen vorbestimmten
Grenzwert überschreitet.
5. Verfahren nach einem der vorhergehenden Ansprüche,
10 dadurch gekennzeichnet,
daß es sich bei dem Buchstabier-Erkennungsmodus um einen Al-
phabetwort-Erkennungsmodus (BM2) handelt.
6. Verfahren nach Anspruch 5,
15 dadurch gekennzeichnet,
daß die Alphabetwörter ("Anton", "Berta", "Cäsar") benutzer-
definierbar sind und/oder daß für einen Buchstaben mehrere
Alphabetwörter verwendet werden können.
- 20 7. Verfahren nach einem der Ansprüche 1 bis 4
dadurch gekennzeichnet,
daß es sich bei dem Buchstabier-Erkennungsmodus um einen
Buchstaben-Erkennungsmodus (BM1) handelt.
- 25 8. Einrichtung zur Spracherkennung eines Wortes,
- mit einem Worterkenner (2, WM), der ein gesprochenes Wort
("Padua") in ein elektrisches Wort-Sprachsignal
(S("Padua")) umsetzt und das Wort-Sprachsignal (S("Padua"))
zur Erkennung des Wortes aus einem Vokabular vorgegebener
30 Wörter (W1, W2) analysiert,
- mit einer Auswahllogik, die, sofern eine eindeutige Zuord-
nung des Wort-Sprachsignals (S("Padua")) zu einem Wort des
Vokabulars nicht getroffen werden kann, eine Vorauswahl von
mehreren Wörtern (W1, W2) des Vokabulars, unter denen sich
35 das gesprochene Wort ("Padua") mit hinreichender Sicherheit
befindet, zusammenstellt,

11

- mit einer Logik (5) zum Bestimmen mindestens eines Entscheidungsbuchstabens zu jedem Wort (W1, W2) des Vokabulars, welcher ein Wort von den anderen Wörtern der Vorauswahl unterscheidbar macht,
 - 5 - mit einer Ausgabeeinheit (7, 8), über die die Entscheidungsbuchstaben zu den Wörtern (W1, W2) der Vorauswahl abgefragt werden, und
 - mit einem Buchstabiererkenner (2, BM1, BM2), der eine Buchstabier-Spracheingabe ("de", "Dora") in ein zugehöriges
 - 10 elektrisches Buchstabier-Sprachsignal (S("de"), S("Dora")) umsetzt und zur Erkennung des Entscheidungsbuchstabens analysiert.
9. Navigationssystem für ein Fortbewegungsmittel, enthaltend
- 15 eine Einrichtung nach Anspruch 8 zur Eingabe von Zielortinformation.

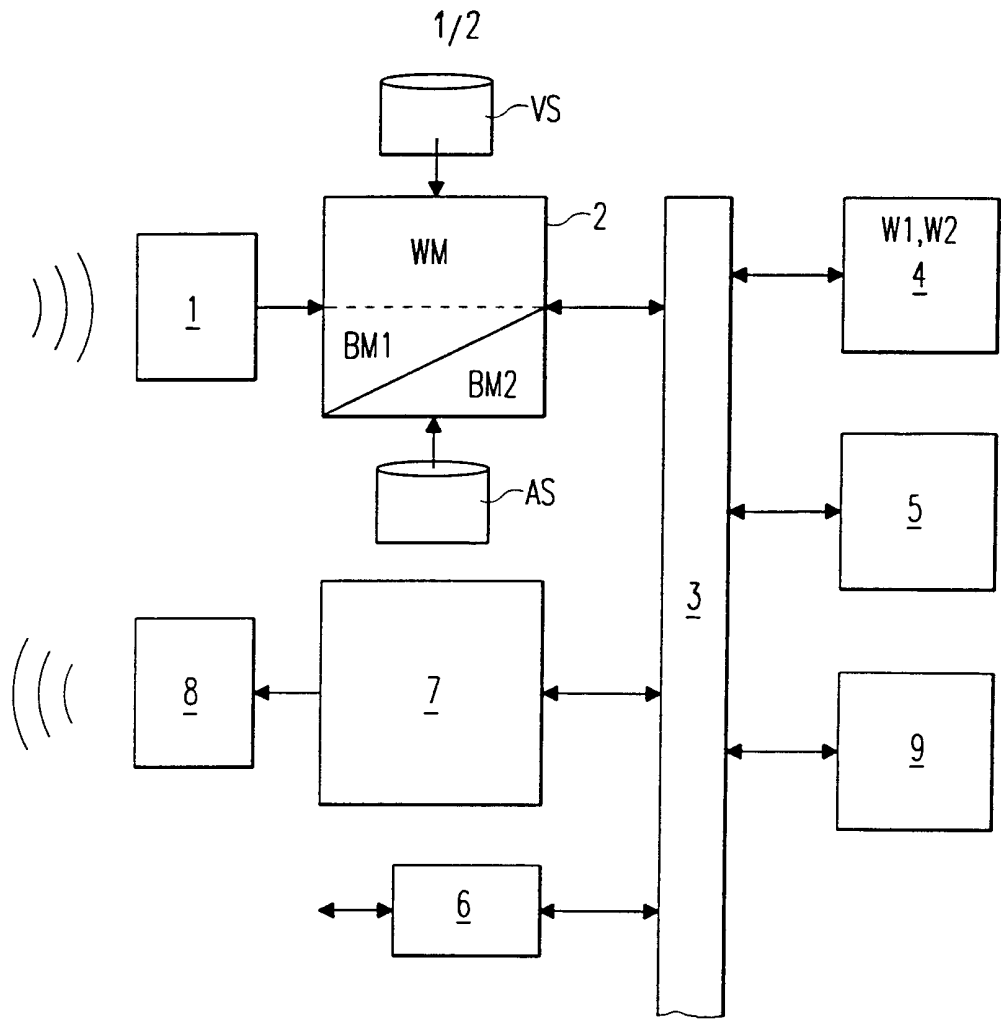


Fig. 1

Typ \ MOD	WM	BM1	BM2
Wort/ Buchstabe	Padua	d	d
gesprochenes Wort/ Buchstabe/Alphabet- wort	"Padua"	"de"	"Doro"
elektrisches Sprachsignal	S("Padua")	S("de")	S("Doro")

Fig. 3

2/2

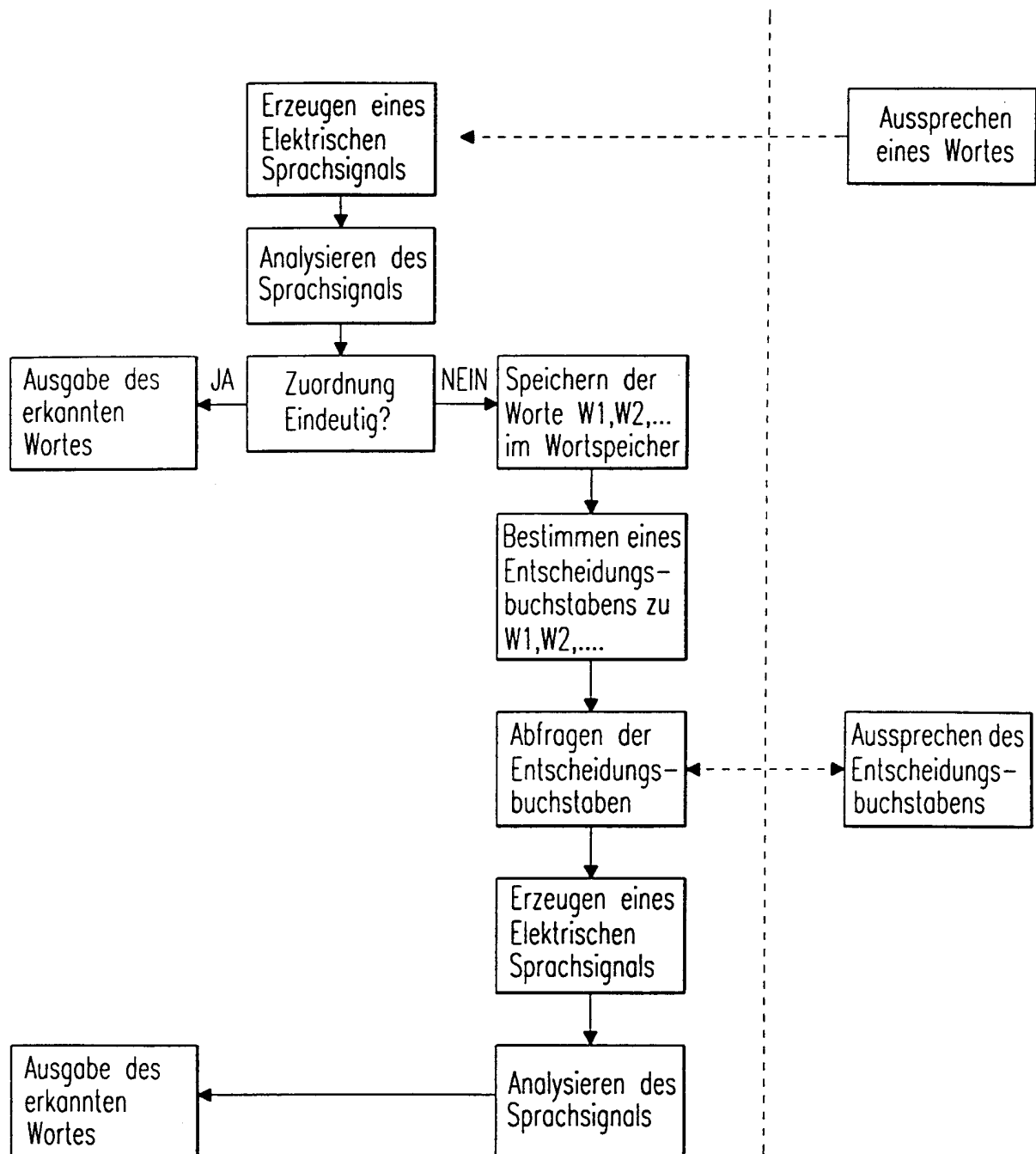


Fig. 2