



(12) Wirtschaftspatent

Erteilt gemäß § 17 Absatz 1 Patentgesetz

(19) DD (11) 218 969 A1

4(51) G 10 L 1/08

AMT FÜR ERFINDUNGS- UND PATENTWESEN

In der vom Anmelder eingereichten Fassung veröffentlicht

(21)	WP G 10 L / 251 185 1	(22)	24.05.83	(44)	20.02.85
------	-----------------------	------	----------	------	----------

(71) siehe (72)

(72) Affa, Karl-Heinz, Dipl.-Ing., 8036 Dresden, Gubener Straße 35, DD

(54) Anordnung zur Bildung von Sprechergruppen

(57) Anordnung zur Bildung von Sprechergruppen, insbesondere für automatische Spracherkennungssysteme, durch die Bildung von Mittelwertfunktionen. Die Aufgabe der Erfindung besteht darin, die zwischen den einzelnen Sprechern bestehenden unterschiedlichen Ähnlichkeitsbeziehungen zu nutzen, um Gruppen von Sprechern zu schaffen. Erfindungsgemäß wird die Aufgabe dadurch gelöst, daß die Sprachsignale jedes Sprechers von einer Signalquelle, einem Mikrophon oder Magnetband über eine Kettenschaltung, bestehend aus einem Vorverstärker, einem Preemphasis-Filter, einer aus einem System von Bandpässen aufgebauten Filterbank, einem Analogmultiplexer und einem Analog-Digital-Umsetzer in Form eines 8 bit Datenwortes an den Digitalrechner übertragen werden. Über einen definierten Zeitraum werden die Amplitudenwerte kanalweise aufsummiert und abschließend auf die Anzahl der betrachteten Zeitfenster normiert. Das so entstandene Mittelwertspektrum wird in einem externen Speicher geladen. Aus den gespeicherten Mittelwertspektren von N untersuchten Sprechern wird eine Ähnlichkeitsmatrix ermittelt und dazu als Ähnlichkeitsmaß die City-Block-Distanz, der Euklidische Abstand oder andere bekannte Abstandsmaße verwendet. Die Sprechergruppen werden in einer folgenden Stufe gebildet. Dazu wird zeilenweise jedes Abstandsmaß in einer weiteren Stufe mit einer Ähnlichkeitsschwelle verglichen. Fig. 4

Anordnung zur Bildung von Sprechergruppen

Anwendungsgebiet der Erfindung

Anordnung zur Bildung von Sprechergruppen, insbesondere für automatische Spracherkennungssysteme, durch die Bildung von Mittelwertfunktionen.

Elemente der Erfindung können in Sprecheridentifikationssystemen und Sprecherverifikationssystemen verwendet werden.

Charakteristik der bekannten technischen Lösungen

Die ersten Forschungen auf dem Gebiet der automatischen Spracherkennung gehen bis in die 50er Jahre zurück. Seit dieser Zeit sind eine Vielzahl von automatischen Erkennungssystemen bekannt geworden, die unter Heranziehung fast ebenso vieler Sprachmerkmale mehr oder weniger zuverlässige Erkennungsergebnisse erzielen.

(HOLMGREN, G.L.: Speaker Recognition, Speech Characteristics, Speech Evaluation and Modification of Speech Signals - A Selected Bibliography
in: IEEE Transactions of Audio and Electroacoustics, 1966, S. 32 - 39).

Dabei tritt die Individualität der menschlichen Sprache von Anfang an als Störfaktor in Erscheinung. Um diesem Sachverhalt Rechnung zu tragen, wurden bisher eine Reihe von Einschränkungen notwendig.

Die gegenwärtig realisierten automatischen Spracherkennungssysteme verwenden vorwiegend Worte als kleinste zu erkennende Einheiten (Einzelworterkenner). Dabei existieren noch Systeme, deren kleinste zu erkennende Einheiten Silben oder

Phoneme darstellen.

Die erwähnten Einschränkungen, die durch die Sprecherindividualität der menschlichen Sprache bedingt sind, führten bei der Entwicklung der Einzelwörterkenner zu zwei Grundrichtungen:

1. Für jeden, das Spracherkennungssystem benutzenden Sprecher wird ein gesonderter Referenzdatensatz gebildet. Dabei ist es notwendig, daß für jeden Sprecher der vollständige Lernzyklus abgearbeitet werden muß. Bevor ein neuer Sprecher an das System herantritt, wird zuerst der für ihn gültige Referenzdatensatz ausgewählt. Diese Referenzdatenwahl kann automatisch (über die automatische Sprechererkennung) erfolgen oder sie kann dem Erkennungssystem über ein Bedienpult mitgeteilt werden. Der Nachteil dieses Verfahrens besteht darin, daß für jeden das System benutzenden Sprecher ein vollständiger Referenzdatensatz ermittelt werden muß. Das erfordert einen relativ großen Aufwand, da der Rechenzeitbedarf für die Anlernphase in der Regel um ein Vielfaches höher liegt als für die Erkennungsphase.

(US-PS 3261916

LINK, W.: Die Erkennung von Sprache durch elektronische Systeme

in: Nachrichten-Elektronik, 1978, Heft 9).

2. Bei diesem Verfahren wird ein gemeinsamer Referenzdatensatz für alle Sprecher gebildet. Um die unterschiedlichen Sprechereigenarten berücksichtigen zu können, ist es deshalb notwendig, für jedes Wort mehrere Referenzmuster zu speichern. Dadurch vergrößert sich der Speicherplatzbedarf notwendigerweise um ein Vielfaches gegenüber dem ersten Verfahren. Um diesen Aufwand zumindest teilweise reduzieren zu können, werden mit Hilfe von Clusterverfahren Repräsentanten von Referenzworten mit gleicher Wortbedeutung gebildet. Der Speicherplatzbedarf bleibt jedoch weiterhin relativ hoch. Dadurch, daß die Anzahl der Referenzworte gegenüber dem ersten Verfahren um ein Vielfaches höher liegt, erhöht sich auch die Erkennungszeit.

(DE-OS 2240557

SCOTT, P.B.: VICI-A Speaker Independent Word Recognition System

in: Int. Conf. on ASSP, 1976, S. 210 - 213).

Die Anzahl der zu verarbeitenden Worte und Sprecher wird durch den zur Verfügung stehenden Speicherplatz für die Speicherung der Referenzdaten bestimmt. Darüber hinaus muß noch die Möglichkeit bestehen, den eigentlichen Erkennungsprozeß in relativ kurzer Zeit (möglichst Echtzeit) durchzuführen. Das bedeutet also, daß ohne wesentliche Vergrößerung des Speicherplatzes eine Erhöhung der Sprecherzahl gleichzeitig eine Verringerung der Zahl der zu erkennenden Worte zur Folge hat. Ebenso zieht eine Erhöhung der Zahl der zu erkennenden Worte notgedrungen eine Verringerung der Sprecherzahl nach sich. Da der in einem Digitalrechner zur Verfügung stehende Speicherplatz derzeit noch sehr begrenzt und darüber hinaus auch noch relativ teuer ist, wirken sich die genannten Einschränkungen ungünstig auf eine breitere Anwendung aus.

Bestrebungen, diese Nachteile durch eine geeignete Merkmalkvektornormierung oder Merkmalkvektortransformation zu kompensieren, stecken noch in den Anfängen und sind bisher von keinem Erfolg gekrönt.

(VINCZJUK, T.K.; KULIAS, A.I.: Sadaca podstrojki, pod diktora pri raspoznavanii reci
in: OBRABOTKA I RASPOZNAVANIE SIGNALOV, Kiew, 1975

RAMISHVILI, G.S.: Recevoj signal i individualnost golosa Isdatjelstwo "Mezniereba", Tbilissi, 1976).

Den bisherigen Lösungen haftet der Mangel an, daß bei konstanter Größe des Hauptspeichers eine Erhöhung der Sprecherzahl nur durch eine Reduzierung der Anzahl der zu erkennenden Worte möglich ist. Zum anderen hat eine Vergrößerung der Zahl der zu erkennenden Worte eine Verringerung der Sprecherzahl zur Folge.

Ziel der Erfindung

Das Ziel der Erfindung ist es, daß ohne eine nennenswerte Vergrößerung des notwendigen Speicherplatzbedarfes des Hauptspeichers eine größere Anzahl von Sprechern mit dem auto-

matischen Spracherkennungssystem arbeiten können.

Darlegung des Wesens der Erfindung

Die Aufgabe der Erfindung besteht darin, die zwischen den einzelnen Sprechern bestehenden unterschiedlichen Ähnlichkeitsbeziehungen zu nutzen, um Gruppen von Sprechern zu schaffen, die sich dadurch auszeichnen, daß sich die Mitglieder einer Sprechergruppe hinsichtlich ihrer zur Spracherkennung verwendeten Sprachmerkmale ähnlich sind.

Erfindungsgemäß wird die Aufgabe dadurch gelöst, daß die Sprachsignale jedes Sprechers von einer Signalquelle, einem Mikrofon oder Magnetband über eine Kettenschaltung, bestehend aus einem Vorverstärker, einem Preemphasis-Filter, einer aus einem System von Bandpässen aufgebauten Filterbank, einem Analogmultiplexer und einem Analog-Digital-Umsetzer in Form eines 8 bit Datenwortes, an den Digitalrechner übertragen werden. Über einen definierten Zeitraum werden die Amplitudenwerte kanalweise aufsummiert und abschließend auf die Anzahl der betrachteten Zeitfenster normiert. Für das so entstandene Mittelwertspektrum von N untersuchten Sprechern wird eine Ähnlichkeitsmatrix ermittelt und dazu als Ähnlichkeitsmaß die City-Block-Distanz, der Euklidische Abstand oder andere bekannte Abstandsmaße verwendet.

Die Sprechergruppen werden in einer folgenden Stufe gebildet. Dazu wird zeilenweise jedes Abstandsmaß in einer weiteren Stufe mit einer Ähnlichkeitsschwelle verglichen. Alle Indizes der Abstandsmaße, welche die Ähnlichkeitsschwelle unterschreiten, werden an eine Stufe zur Ausgabe der Sprechernummern der ermittelten Sprechergruppen weitergegeben, in aufsteigender Folge sortiert und ausgegeben. In automatischen Spracherkennungssystemen erfolgt in einer weiteren Stufe die Bildung eines Mittelwertspektrums zur Vorauswahl des zur Klassifikation verwendeten Referenzdatensatzes. In einem externen Speicher ist dazu eine Auswahl von Referenzdatensätzen unterschiedlicher Sprechertypen gespeichert. In die Mittelwertbildung werden alle vorher gesprochenen Erkennungseinheiten einbezogen. Nach jedem gesprochenen Wort wird überprüft, ob die getroffene Referenzdatenauswahl noch gültig ist. Ein neuer Re-

ferenzdatensatz wird in den Hauptspeicher geladen, wenn dieser die Ähnlichkeitsbedingungen besser erfüllt.

Zur Bildung von Sprechergruppen werden als Sprachmerkmale Merkmalvektorfolgen von Amplitudenhistogrammen, Nulldurchgangsfunktionen und LPC-Analysefunktionen verwendet.

Bei der Benutzung eines automatischen Spracherkennungssystems durch einen Fremdsprecher wird aus vorliegenden Sätzen von Referenzdaten derjenige Datensatz zur Erkennung herangezogen, der mit hoher Sicherheit das beste Erkennungsergebnis liefert.

Ausführungsbeispiel

Die Erfindung soll nachstehend an einem Ausführungsbeispiel näher erläutert werden.

In der zugehörigen Zeichnung zeigen:

Fig. 1: Blockschaltbild der Anordnung zur Bildung von Sprechergruppen

Fig. 2: Mittelwertspektrum eines weiblichen und eines männlichen Sprechers

Fig. 3: Prinzip der Sprechergruppenbildung allgemein

Fig. 4: Prinzip der Sprechergruppenbildung in einem automatischen Spracherkennungssystem

Den Ausgangspunkt für die Bildung von Sprechergruppen bilden Mittelwertfunktionen von verschiedenen Sprechern. Diese Mittelwertfunktionen werden durch komponentenweise Mittelung von Parameterfunktionen gebildet. Dabei müssen die Parameterfunktionen als Vektorfolgen konstanter Komponentenzahl darstellbar sein. Diese Bedingungen erfüllen z. B. Filterspektren, Nulldurchgangsfunktionen, Amplituden- bzw. Pegelhistogramme und LPC-Koeffizienten.

Die Verwendung von Filterspektren zur automatischen Spracherkennung ist sehr weit verbreitet. In diesem Fall werden Mittelwertspektren zur Ermittlung der Sprechergruppen herangezogen. In Fig. 1 wird eine Anordnung gezeigt, mit der Mittelwertspektren unterschiedlicher Sprecher gebildet werden können. Die Sprachsignale jedes Sprechers gelangen von einer Signalquelle 1 (Mikrofon, Magnetband o. dgl.) über einen Vor-

verstärker 2 auf ein Preemphasis-Filter 3. Dieses Filter bewirkt eine Höhenanhebung des Sprachsignals um etwa 6 - 12 dB pro Oktave, um dem bei der Sprache typischen Amplitudenabfall zu hohen Frequenzen hin entgegenzuwirken. Danach gelangt das Sprachsignal zur Parallelfilterbank 4. Diese Parallelfilterbank besteht aus einem System von Bandpässen. Die Untersuchungen wurden mit einer 20kanaligen Filterbank durchgeführt, die die nachfolgend angegebenen Grenzfrequenzen und Mittelfrequenzen besitzen.

Kanalnummer	Mittenfrequenz (Hz)	Bandbreite (Hz)
1	109	70 - 170
2	214	170 - 270
3	316	270 - 370
4	417	370 - 470
5	518	470 - 570
6	623	570 - 680
7	742	680 - 810
8	886	810 - 970
9	1065	970 - 1170
10	1280	1170 - 1400
11	1535	1400 - 1680
12	1833	1680 - 2000
13	2191	2000 - 2400
14	2629	2400 - 2880
15	3157	2880 - 3460
16	3789	3460 - 4150
17	4546	4150 - 4980
18	5457	4980 - 5980
19	6548	5980 - 7170
20	7853	7170 - 8600

Ein sich daran anschließender Analogmultiplexer 5 tastet alle 12,5 ms die Filterausgänge schrittweise ab. Die so entstandene analoge Funktion wird vom Analog-Digital-Umsetzer 6 in ein 8 bit langes Datenwort umgewandelt. Diese 8 bit lassen eine Auflösung von 128 Amplitudenstufen zu. Das entspricht einem Dynamikumfang von 48 dB. Die weitere Verarbeitung erfolgt rechentechnisch. Das digitalisierte Signal gelangt in den Rechner 7 (KRS 4201) und wird dort gemittelt.

den.

Die City-Block-Distanz läßt sich nach der Beziehung

$$CD_{ij} = \sum_{l=1}^L |x_i(l) - x_j(l)|$$

und der Euklidische Abstand nach der Beziehung

$$ED_{ij} = \sum_{l=1}^L \sqrt{(x_i(l) - x_j(l))^2}$$

ermitteln. Dabei bezeichnet i und j das Sprecherpaar, für das der Abstand ermittelt wird. Die Komponente der Parameterfunktion wird durch l angegeben. L ist die Anzahl der Komponenten, aus denen sich der Parametervektor x zusammensetzt.

In der Stufe 13 erfolgt die Bildung der Sprechergruppen. Dabei wird so vorgegangen, daß zeilenweise jedes Abstandsmaß mit einer Ähnlichkeitsschwelle Stufe 12 verglichen wird. Die Ähnlichkeitsschwelle wird experimentell bestimmt. Sie sollte so gewählt werden, daß etwa 2 - 5 Sprecher einer Gruppe zugeordnet werden. Nachdem die gesamte Zeile der Ähnlichkeitsmatrix abgearbeitet wurde, werden alle Indizes der Abstandsmaße, die die Ähnlichkeitsschwelle unterschreiten, an die Stufe 14 weitergegeben. Die Indizes der Abstandsmaße stimmen mit den Nummern der betrachteten Sprecher überein. In der Stufe 14 werden die Sprechernummern, die nach Abarbeitung jeder Zeile der Ähnlichkeitsmatrix übermittelt werden, in aufsteigender Folge sortiert. Die so entstandenen Folgen werden verglichen, um Wiederholungen auszuschließen. Abschließend werden die Sprechernummern jeder Sprechergruppe ausgegeben.

Fig. 4 zeigt die Anordnung zur Sprechergruppenbildung in einem automatischen Spracherkennungssystem. Ein wesentlicher Vorteil des Systems ist die dynamische Sprecheranpassung. Dabei erfolgt eine ständige Aktualisierung der Referenzdaten. Für die Eingabe der Sprachdaten gilt das in bezug auf Fig. 1 Gesagte. Dies betrifft die Stufen 1 - 6.

Während innerhalb des Rechners die eingegebene Vektorfolge zum einen zur Klassifikation verwendet wird, erfolgt zum anderen die Bildung eines Mittelwertspektrums zur Vorauswahl

des zur Klassifikation verwendeten Referenzdatensatzes in einer Stufe 15. Auf einem externen Speicher 16 (Digitalmagnetband, Digitalkassette, Magnetplatte o. dgl.) sind in geeigneter Weise eine Auswahl von Referenzdatensätzen unterschiedlicher Sprechertypen gespeichert. Mit Hilfe des in Stufe 15 gebildeten Mittelwertspektrums wird der Referenzdatensatz ausgewählt, dessen Mittelwertspektrum (über den gesamten Referenzdatensatz gebildet) die größte Ähnlichkeit zu diesem neu ermittelten Mittelwertspektrum besitzt. Dabei gehen alle vorher gesprochenen Erkennungseinheiten in die Mittelung ein. Nach jedem gesprochenen Wort wird überprüft, ob die getroffene Referenzdatenauswahl 17 noch gültig ist. Im anderen Fall wird ein anderer Referenzdatensatz, der die Ähnlichkeitsbedingung besser erfüllt, in den Hauptspeicher geladen. Dadurch wird erreicht, daß bei Sprecherwechsel eine Anpassung des Erkennungssystems an den neuen Sprecher erreicht wird. Da die meisten Spracherkennungssysteme die Möglichkeit vorsehen, ein falsch erkanntes Wort Stufe 18 zurückzuweisen, ist es sogar möglich, den ausgewählten Referenzdatensatz ständig zu aktualisieren. Immer wenn eine Erkennungseinheit richtig erkannt wurde, erfolgt die Ermittlung eines neuen Referenzdatenwortes, in dem der Mittelwert zwischen dem eben gesprochenen Text und dem bisherigen Referenzdatenwort berechnet wird Stufe 19. Damit wird erreicht, daß die während des normalen Sprechablaufs typischen Sprachveränderungen eines Sprechers Berücksichtigung finden. In der Stufe 20 erfolgt die durch die Spracheingabe bezweckte Reaktion des Erkennungssystems.

Die Tabelle 1 enthält die Erkennungsergebnisse von 10 männlichen Sprechern. Es wurden insgesamt 30 unterschiedliche Worte verwendet, mit jeweils 7 Realisierungen pro Wort. Das ergibt einen Gesamtumfang von 210 Worten pro Sprecher.

Mit den in den Figuren 1 und 3 beschriebenen Anordnungen wurde die Ähnlichkeitsmatrix für diese Sprecher ermittelt. Als Abstandsmaß wurde die City-Block-Distanz verwendet. Die Ähnlichkeitsmatrix ist in Tabelle 2 enthalten. Die Ähnlichkeitschwelle zur Sprechergruppenbildung wurde so gewählt, daß Gruppen von maximal zwei Sprechern gebildet werden. Für größere Sprechergruppen ist eine höhere Sprecherzahl nötig. Die

Tabelle 3 enthält die Erkennungsergebnisse für zwei Sprechergruppen (Gruppe 1 aus Sprecher 2 und 8, Gruppe 2 aus Sprecher 5 und 7 gebildet).

Die Erkennungsergebnisse der Sprecher innerhalb einer Sprechergruppe liegen über 90 %. Beim Erkennungsvorgang befindet sich jeweils nur der Referenzdatensatz für eine Sprechergruppe im Hauptspeicher. Welcher Referenzdatensatz das ist, wird während des laufenden Erkennungsprozesses entschieden. Neben den Referenzsprachdaten der jeweiligen Sprechergruppe sind außerdem die Mittelwertspektren sämtlicher zur Verfügung stehenden Referenzdatensätze der unterschiedlichen Sprechergruppen gespeichert.

Parallel zum Erkennungsprozeß wird das Mittelwertspektrum des Nutzers ermittelt. Dabei gehen vorherliegende Sprachäußerungen in die Mittelung ein. Es wird bei jedem Erkennungsprozeß geprüft, zu welchem Mittelwertspektrum der Referenzdatensätze der Sprechergruppen die größte Ähnlichkeit besteht. Dieser Referenzdatensatz wird von einem Externspeicher (Magnetband, Magnetplatte, Magnetkassette o. ä.) in den Hauptspeicher geladen. Damit arbeitet das Erkennungssystem gewissermaßen sprecheranpassend. Bei Sprecherwechsel kann automatisch auf einen geeigneten Referenzdatensatz zurückgegriffen werden.

Anhand der Tabellen 1 und 2 soll verdeutlicht werden, wie bei Nutzung des automatischen Spracherkennungssystems durch einen Fremdsprecher verfahren wird. Als Referenzdatensätze stehen zum Beispiel Muster der Sprecher 1 - 5 zur Verfügung. Im allgemeinen Fall können das auch Referenzdatensätze von Sprechergruppen sein. Die Sprecher 6 - 10 stellen Fremdsprecher dar. Damit gelten die ersten 5 Zeilen der Tabelle 1 und 2.

Aus Tabelle 2 ist ersichtlich, daß zum Beispiel für Sprecher 10 (Fremdsprecher) der Sprecher 3 das geringste Abstandsmaß (0,81) liefert. Tabelle 1 zeigt, daß unter Verwendung des Referenzdatensatzes des Sprechers 3 für den Sprecher 10 das höchste Erkennungsergebnis (86,3 %) erzielt wird.

Für Sprecher 8 liefert der Referenzdatensatz des Sprechers 2 (Abstandsmaß 0,73) das höchste Erkennungsergebnis (87,6 %). Für Sprecher 7 wird demzufolge mit dem Referenzdatensatz des

Sprechers 5 (Abstandsmaß 0,78) das beste Erkennungsergebnis (90,4 %) erreicht.

Nutzer

		1	2	3	4	5	6	7	8	9	10
angelernter Sprecher	1	96,2	81,8	61,9	83,8	91,9	85,2	78,0	83,3	54,7	50,4
	2	81,9	99,0	78,1	77,4	84,4	79,5	79,4	87,6	72,4	70,0
	3	74,3	74,2	99,0	82,4	85,2	72,9	74,2	86,4	67,5	86,3
	4	71,9	72,2	65,2	97,1	80,0	78,1	76,1	72,4	69,0	71,4
	5	79,5	80,3	77,0	78,6	99,0	71,4	90,4	62,5	66,0	62,4
	6	72,1	71,2	76,3	82,2	78,3	97,8	69,8	72,9	71,8	78,3
	7	60,9	65,3	63,2	79,1	86,3	67,3	98,1	72,1	66,3	62,1
	8	63,2	75,3	69,1	68,2	62,1	64,1	62,7	97,3	73,2	77,3
	9	59,8	62,3	60,3	69,8	72,1	67,3	69,1	73,2	97,8	80,3
	10	57,9	53,8	49,7	65,2	68,3	79,2	62,7	85,1	84,2	98,5

Tabelle 1 (Angaben in %)

Sprecher

		1	2	3	4	5	6	7	8	9	10
Sprecher	1	0,00	1,23	1,69	1,19	1,22	1,41	1,39	1,10	1,11	1,26
	2	1,23	0,00	1,16	1,02	0,93	1,32	1,02	0,73	1,22	1,33
	3	1,69	1,16	0,00	1,06	0,99	1,00	1,34	0,93	1,14	0,81
	4	1,19	1,02	1,06	0,00	0,81	0,84	0,98	0,93	1,08	0,99
	5	1,22	0,93	0,99	0,81	0,00	1,03	0,78	1,02	1,09	1,12
	6	1,41	1,32	1,00	0,84	1,03	0,00	1,28	1,19	1,12	0,95
	7	1,39	1,02	1,34	0,98	0,78	1,28	0,00	1,31	1,36	1,47
	8	1,10	0,73	0,93	0,93	1,02	1,19	1,31	0,00	0,87	0,78
	9	1,11	1,22	1,14	1,08	1,09	1,12	1,36	0,87	0,00	0,95
	10	1,26	1,33	0,81	0,99	1,12	0,95	1,47	0,78	0,95	0,00

Tabelle 2 (Angaben ohne Maßeinheit (normiert))

Nutzer

angelernter Sprecher	Nutzer									
	1	2	3	4	5	6	7	8	9	10
2,8	83,8	95,7	82,9	81,9	84,3	79,5	84,2	90,4	81,5	75,7
5,7	67,7	81,3	76,2	87,6	97,1	63,3	94,7	68,9	72,4	71,9

Tabelle 3 (Angaben in %)

Erfindungsanspruch

1. Anordnung zur Bildung von Sprechergruppen, insbesondere für automatische Spracherkennungssysteme, durch die Bildung von Mittelwertfunktionen dadurch gekennzeichnet, daß die Sprachsignale jedes Sprechers von einer Signalquelle (1), einem Mikrofon oder Magnetband, über eine Ketten-schaltung, bestehend aus einem Vorverstärker (2), einem Preemphasis-Filter (3), einer aus einem System von Band-pässen aufgebauten Filterbank (4), einem Analogmultiple-xer (5) und einem Analog-Digital-Umsetzer (6), in Form eines 8 bit Datenwortes an den Digitalrechner (7) über-tragen werden, über einen definierten Zeitraum die Ampli-tudenwerte kanalweise aufsummiert und abschließend auf die betrachteten Zeitfenster normiert werden, das entstandene Mittelwertspektrum in einen externen Speicher (8) geladen und aus den gespeicherten Mittelwertspektren (10) von N untersuchten Sprechern eine Ähnlichkeitsmatrix (11) er-mittelt und dazu als Ähnlichkeitsmaß die City-Block-Di-stanz benutzt wird, in einer folgenden Stufe (12) mit einer Ähnlichkeitsschwelle verglichen, alle Indizes der Abstandsmaße, welche die Ähnlichkeitsschwelle unterschrei-ten, werden an eine Stufe (14) zur Ausgabe der Sprecher-nummern der ermittelten Sprechergruppen weitergegeben, in aufsteigender Folge sortiert und ausgegeben, weiter er-folgt in automatischen Spracherkennungssystemen die Mit-telwertbildung zur Vorauswahl des zur Klassifikation ver-wendeten Referenzdatensatzes in einer Stufe (15), eine Auswahl von Referenzdatensätzen unterschiedlicher Spre-cherarten ist in einem externen Speicher (16) gespeichert, in die Mittelwertbildung werden alle vorher gesprochenen Erkennungseinheiten einbezogen, nach jedem gesprochenen Wort wird überprüft, ob die getroffene Referenzdatenaus-wahl (17) noch gültig ist, und ein neuer Referenzdatensatz wird in den Hauptspeicher geladen, wenn dieser die Ähn-lichkeitsbedingungen besser erfüllt.
2. Anordnung zur Bildung von Sprechergruppen nach Punkt 1 dadurch gekennzeichnet, daß als Sprachmerkmale Merkmal-vektorfolgen von Amplitudenhistogrammen, Nulldurchgangs-funktionen und LPC-Analysefunktionen verwendet werden.

3. Anordnung zur Bildung von Sprechergruppen nach Punkt 1 und 2 dadurch gekennzeichnet, daß als Ähnlichkeitsmaß der Euklidische Abstand, das Kreuzkorrelationsmaß und die Mahalanobis-Distanz verwendet werden.
4. Anordnung zur Bildung von Sprechergruppen nach Punkt 1 dadurch gekennzeichnet, daß bei Benutzung eines automatischen Spracherkennungssystems durch einen Fremdsprecher aus vorliegenden Sätzen von Referenzdaten derjenige Datensatz zur Erkennung herangezogen wird, der mit hoher Sicherheit das beste Erkennungsergebnis liefert.

Hierzu 3 Blatt Zeichnungen.

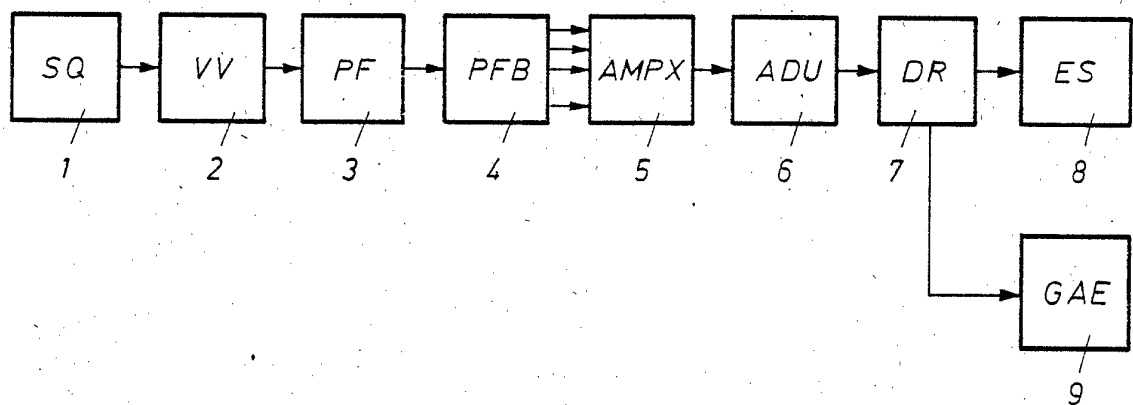


Fig. 1

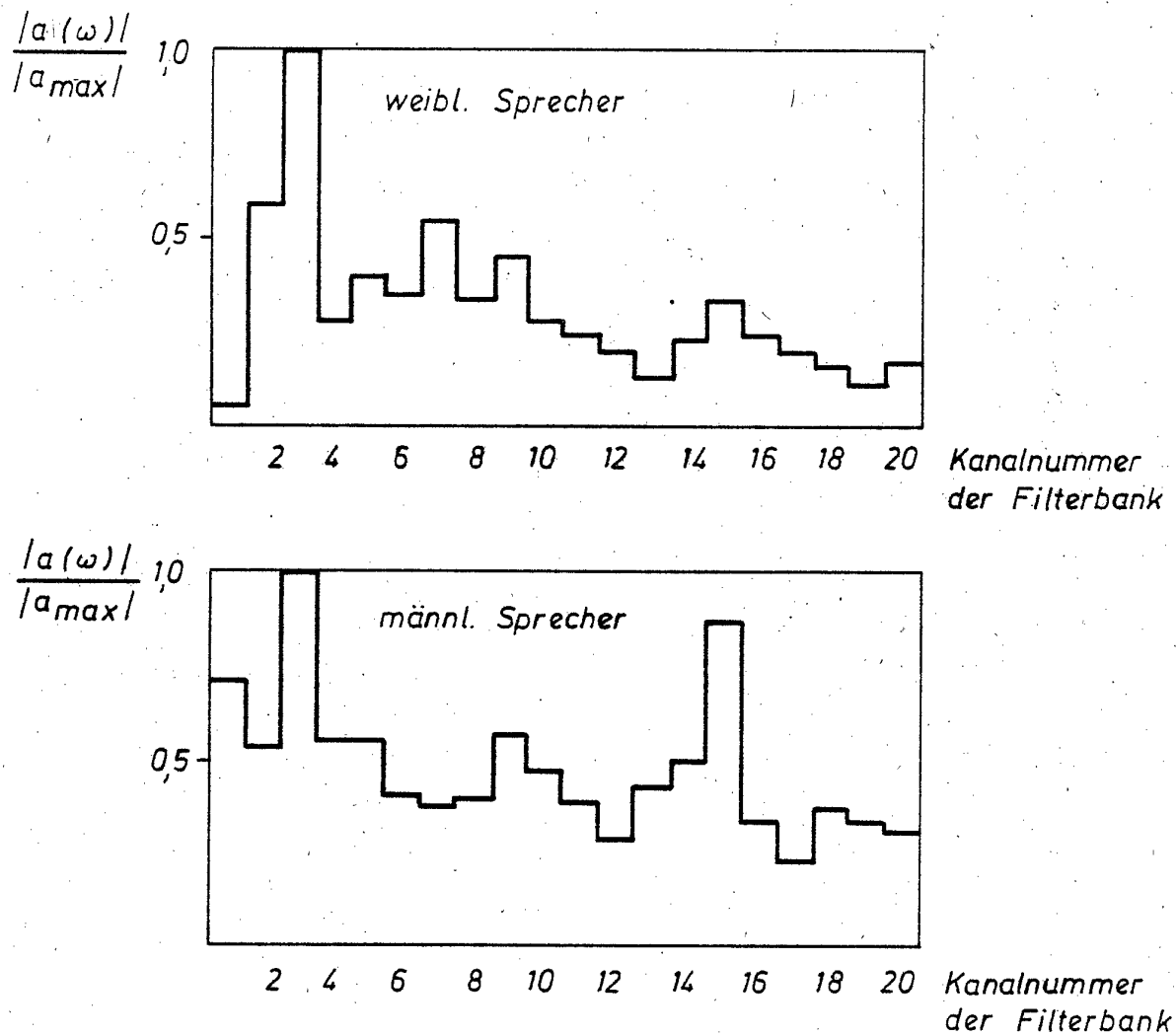


Fig. 2

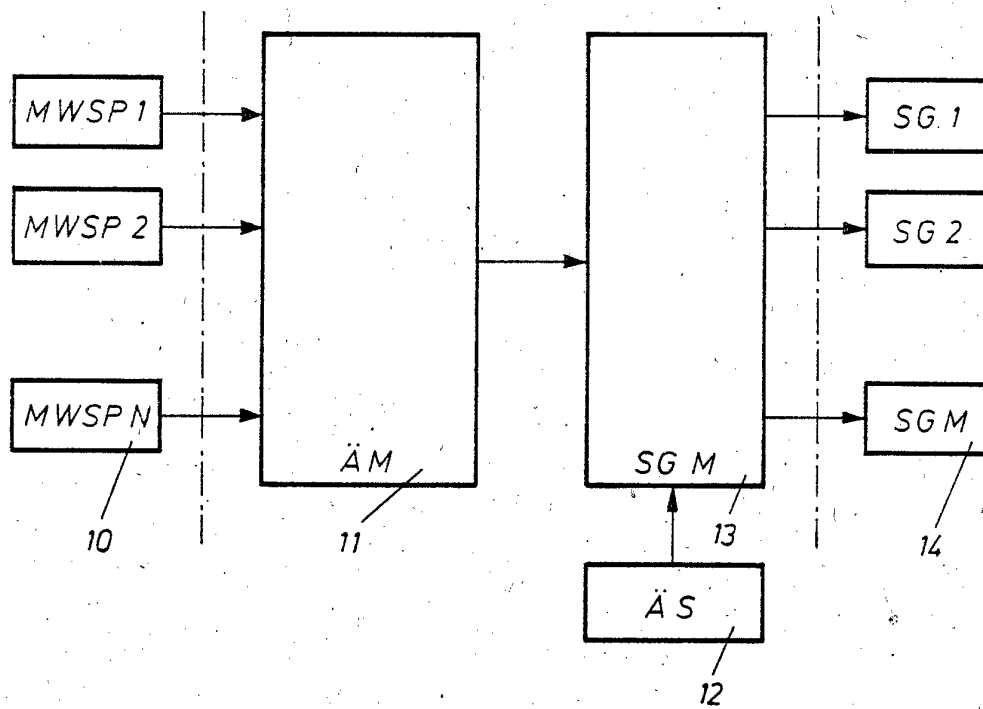


Fig. 3

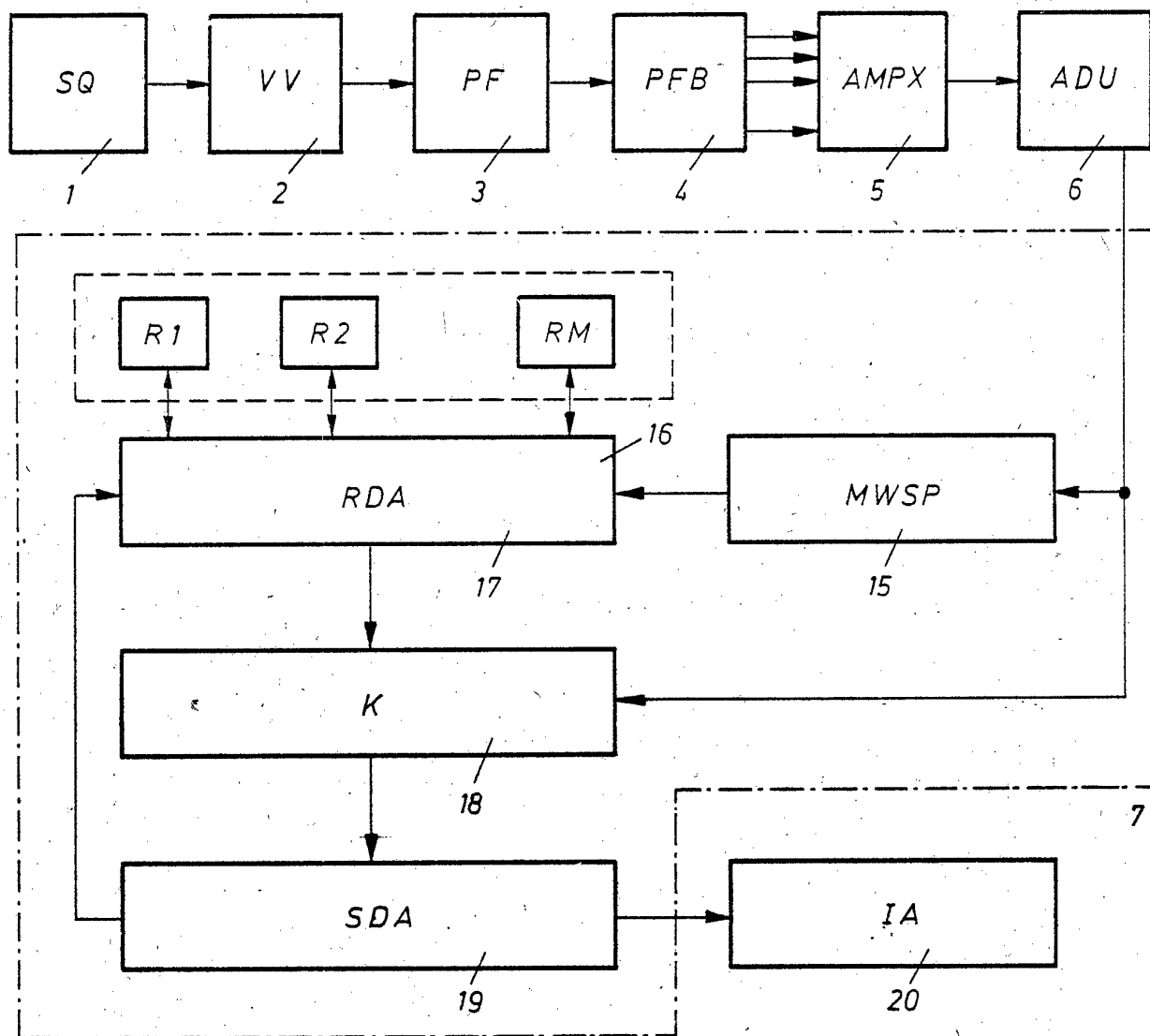


Fig. 4