



Europäisches Patentamt
 European Patent Office
 Office européen des brevets



(11) **EP 1 187 095 A2**

(12) **EUROPÄISCHE PATENTANMELDUNG**

(43) Veröffentlichungstag:
13.03.2002 Patentblatt 2002/11

(51) Int Cl.7: **G10L 13/08**

(21) Anmeldenummer: **01120155.5**

(22) Anmeldetag: **22.08.2001**

(84) Benannte Vertragsstaaten:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
 MC NL PT SE TR**
 Benannte Erstreckungsstaaten:
AL LT LV MK RO SI

(71) Anmelder: **SIEMENS AKTIENGESELLSCHAFT
 80333 München (DE)**

(72) Erfinder: **Hain, Horst-Udo
 81825 München (DE)**

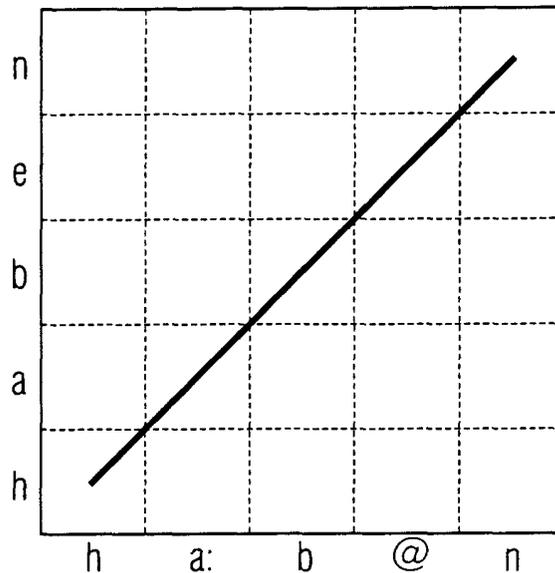
(30) Priorität: **31.08.2000 DE 10042943**

(54) **Graphem-Phonem-Zuordnung**

(57) Das Zuordnen von Phonemen zu den sie erzeugenden Graphemen in einem Lexikon mit Wörtern (Graphemfolgen) und deren zugehöriger Lautschrift (Phonemfolgen) für die Vorbereitung von Mustern für

das Training von Neuronalen Netzen zur Graphem-Phonem-Konvertierung wird mit einer Variante der dynamischen Programmierung durchgeführt, die auch als dynamic time warping (DTW, dynamische Zeitanpassung) bekannt ist.

FIG 2



EP 1 187 095 A2

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren, ein Computerprogrammprodukt, einen Datenträger und ein Computersystem zum Zuordnen von Phonemen zu den sie erzeugenden Graphemen in einem Lexikon mit Wörtern (Graphemfolgen) und deren zugehöriger Lautschrift (Phonemfolgen).

[0002] Sprachverarbeitungsverfahren sind beispielsweise aus US 6 029 135, US 5 732 388, DE 19636739 C1 und DE 19719381 C1 bekannt. Für automatisches Vorlesen oder die Erweiterung des Wortschatzes von Diktiersystemen bzw. von automatischen Spracherkennungssystemen werden Routinen zur Graphem-Phonem-Konvertierung benötigt, d.h. zur Umwandlung geschriebener Wörter in gesprochene Laute. Dazu werden häufig Neuronale Netze eingesetzt.

[0003] Das Training dieser Neuronalen Netze erfolgt mit Hilfe von Mustern. Ein Muster besteht aus einer Anzahl von Buchstaben aus einem Wort, mit denen die Eingangsknoten des Neuronalen Netzes beaufschlagt werden, und dem zugehörigen Phonem, entsprechend dem Ausgangsknoten. Häufig wird jedem Phonem auch noch ein sog. Gruppierungswert zugeordnet. Der Gruppierungswert gibt die Anzahl der Grapheme an, die das zugehörige Phonem erzeugen.

[0004] Die Muster werden aus sogenannten Trainingslexika gewonnen. Ein Trainingslexikon enthält Zuordnungen von Graphemen, in der Regel Wörter, Ziffern etc., also all dem, was konvertiert werden soll, zu Phonemen und Phonemfolgen, also Graphem-Phonem-Transkriptionen auf der Ebene von Wörtern. Die Phonemfolgen werden im Trainingslexikon durch eine geeignete Art von Lautschrift wiedergegeben. Im Bereich der automatischen Spracherkennung werden häufig die Lautschriften SAMPA oder Spicos-Inventar verwendet, die auf ASCII-Zeichen beruhen. Beispielfhaft seien einige deutsche Wörter aufgeführt mit der zugehörigen Lautschrift in SAMPA:

Quatsch	kv'atS
spät	SpE:t
Schutz	SUts
schwer	Sve:6
Sprache	Spra:x@

[0005] Der Laut "sch" wird z.B. durch [S] dargestellt, Dehnungen durch einen Doppelpunkt. Dabei werden Phoneme in eckigen Klammern [] dargestellt, Grapheme im spitzen Klammern <>. Sämtliche Lautschrift-Beispiele der Beschreibung werden in SAMPA wiedergegeben.

[0006] Diese Trainingslexika enthalten zwar die Lautschrift, jedoch nicht die eindeutige Zuordnung von Phonemen und die sie erzeugenden Grapheme, wie sie für die Muster notwendig sind. Zum Beispiel für das Wort <Sprache> wäre die folgende Zuordnung wünschenswert:

Grapheme	S	p	r	a	c	h	e
Phoneme	S, 1	p, 1	r, 1	a:, 1	x, 2		@, 1

aus der sich leicht die Muster für das Training des Neuronalen Netzes ableiten lassen. Bei einem Eingangsfenster mit 7 Buchstaben ergeben sich aus der eindeutigen Zuordnung unmittelbar die folgenden 6 Muster:

1. Muster	Eingang				S	p	r	a
	Ausgang	S,1						

[0007] Der Graphemfolge aus 3 Leerzeichen, <S>, <p>, <r> und <a>, wobei <S> sich zentral im Eingangsfenster befindet, wird der Laut [S] mit dem Gruppierungswert 1 zugeordnet. Entsprechend erhält man als weitere Muster:

2. Muster	Eingang			S	p	r	a	c
	Ausgang	p, 1						

3. Muster	Eingang		S	p	r	a	c	h
	Ausgang	r, 1						

EP 1 187 095 A2

4. Muster	Eingang	S	p	r	a	c	h	e
	Ausgang	a:, 1						

5

5. Muster	Eingang	p	r	a	c	h	e	
	Ausgang	x, 2						

10 **[0008]** Dem Ach-Laut oder stimmlosen velaren Frikativ "ch" ist gemäß den Segmentierungsregeln ein Gruppierungswert von 2 zugeordnet, da ihm die beiden Buchstaben <c> und <h> zugeordnet sind. Daher kann im folgenden Muster das Buchstabenfenster um 2 Buchstaben verschoben werden:

6. Muster	Eingang	a	c	h	e			
	Ausgang	@, 1						

15

20 **[0009]** Die Zuordnung von Buchstaben zu Phonemen ergibt sich aus der Lautschrift des Lexikons jedoch nicht eindeutig. Das Wort <Sprache> besteht aus 7 Buchstaben, aber nur aus 6 Phonemen. Es stellt sich die Frage, welches der Phoneme durch 2 Buchstaben erzeugt wird. Da auch 2 Phoneme durch einen Buchstaben erzeugt werden können, z.B. [ks] durch <x>, ist die Unsicherheit in der Graphem-Phonem-Zuordnung für die Muster ein allgemeines Problem.

[0010] Bisher wurde die Graphem-Phonem-Zuordnung halbautomatisch durchgeführt, ausgehend von für einen Muttersprachler evidenten Erfahrungsregeln, was jedoch insbesondere beim mehrsprachigen Systemen fehleranfällig ist und einen erheblichen Aufwand darstellt.

25 **[0011]** Aufgabe der Erfindung ist es, die Zuordnung von Phonemen zu den sie erzeugenden Graphemen für Muster für das Training eines Neuronalen Netzes für die Graphem-Phonem-Konvertierung automatisch zu erzeugen.

[0012] Diese Aufgabe wird erfindungsgemäß durch ein Verfahren, ein Computerprogrammprodukt, einen Datenträger und ein Computersystem gemäß den unabhängigen Ansprüchen gelöst.

30 **[0013]** Unter einem Computerprogrammprodukt wird dabei das Computerprogramm als handelbares Produkt verstanden, in welcher Form auch immer, z.B. auf Papier, auf einem computerlesbaren Datenträger, über ein Netz verteilt etc.

[0014] Erfindungsgemäß wird das Zuordnen von Phonemen zu den sie erzeugenden Graphemen in einem Lexikon mit Wörtern (Graphemfolgen) und deren zugehöriger Lautschrift (Phonemfolgen) mit Hilfe eines Algorithmus der dynamischen Zeitanpassung (DTW, Dynamic Time Warping) durchgeführt.

35 **[0015]** DTW-Algorithmen sind eine Variante der dynamische Programmierung. Sie sind z.B. beschrieben in:

1. Hoffmann, R.: "Signalanalyse und -erkennung." Springer Verlag, Berlin, Heidelberg, 1998, S. 390 - 393.
2. Rabiner, L.R.; Juang, B.-H.: "Fundamentals of speech recognition." Englewood Cliffs: Prentice Hall 1993 (Prentice Hall Signal Processing Series).
3. Besling, S.: "Heuristical and Statistical methods of Grapheme-to-Phoneme Conversion"; Proceedings KONVENS 94, Wien, S. 23 - 31.

40

45 **[0016]** Vorzugsweise werden in einem ersten Schritt Wörter ausgewählt, bei denen die Anzahl der Grapheme und die Anzahl der Phoneme übereinstimmen. Bei diesen Wörtern werden die Grapheme und Phoneme in der Reihenfolge der Angabe ihrer Grapheme und Phoneme im Lexikon einander zugeordnet. Aus diesen Zuordnungen wird die relative Häufigkeit ermittelt, mit der ein Phonem durch ein Graphem erzeugt wird. Alternativ kann auch die relative Häufigkeit bestimmt werden, mit der ein Graphem einem Phonem zugeordnet wird.

50 **[0017]** In einem zweiten Schritt wird für jedes Wort des Lexikons eine zweidimensionale Matrix erstellt, die sog. Inzidenzmatrix, deren einer Index durch die Grapheme des Worts und deren zweiter Index durch die Phoneme des Worts gegeben ist. Als Einträge der Matrix werden die zu dem jeweiligen Phonem-Graphem-Paar gehörenden, im ersten Schritt bestimmten, relativen Häufigkeiten gewählt.

55 **[0018]** In einem dritten Schritt wird jeder Matrixeintrag durch eine mathematische Operation, insbesondere eine Multiplikation, mit dem Extremalwert, der vorzugsweise der Maximalwert ist, der folgenden drei vorhergehenden Matrixeinträge verknüpft: dem Eintrag für dasselbe Phonem und das vorhergehende Graphem im Wort, dem Eintrag für das vorhergehende Phonem und dasselbe Graphem im Wort und dem Eintrag für das vorhergehende Phonem und das vorhergehende Graphem im Wort. Statt der Multiplikation sind auch andere Rechenoperationen denkbar, etwa eine Addition der Kehrwerte der Matrixeinträge oder andere, in der dynamischen Programmierung bewährte Opera-

tionen.

[0019] Bei der Multiplikation wird mit dem ersten Graphem und dem ersten Phonem des Worts begonnen, wobei bei der Bestimmung der Maximalwerte die sich jeweils aus den Multiplikationen ergebenden geänderten Einträge der Matrix herangezogen werden. Durch das Bestimmen, welcher der drei vorhergehenden Matrixeinträge eines Matrixeintrags maximal war, wird eine Schrittrichtung für diesen Matrixeintrag ermittelt.

[0020] In einem vierten Schritt wird ausgehend vom Matrixeintrag für das letzte Phonem und das letzte Graphem jeweils die für den Matrixeintrag ermittelte Schrittrichtung entlang ein Weg durch die Matrix definiert bis zum Matrixeintrag für das erste Phonem und das erste Graphem. Die zu dem Weg gehörenden Matrixelemente definieren die Zuordnung von Graphemen zu Phonemen des Worts.

[0021] Das Lexikon ist danach konsistent aufbereitet. Das erfindungsgemäße Verfahren kann für das Erzeugen von Mustern zum Training von Neuronalen Netzen angepasst werden.

[0022] Nach Ausführen der Zuordnung von Graphemen zu Phonemen für jedes Wort des Lexikons, wird aus diesen Zuordnungen die positionsabhängige relative Häufigkeit ermittelt, mit der ein Phonem durch zwei oder mehr Grapheme erzeugt wird oder zwei oder mehr Phoneme durch ein Graphem erzeugt werden oder zwei oder mehr Grapheme einem Phonem zugeordnet werden oder ein Graphem zwei oder mehr Phonemen zugeordnet wird. Dies ermöglicht es, in einem weiteren Schritt Korrekturen an den Zuordnungen vorzunehmen.

[0023] Diese korrigierten Zuordnungen können zum iterativen Verbessern der relativen Häufigkeiten und damit der Zuordnungen eingesetzt werden. Dazu werden nach dem Korrigieren der Zuordnungen für jedes Wort des Lexikons aus diesen korrigierten Zuordnungen erneut die positionsabhängigen relativen Häufigkeiten ermittelt. Diese werden bei weiteren Zuordnungen verwendet.

[0024] Beim Bestimmen der relativen Häufigkeiten ist es vorteilhaft, nur solche Zuordnungen zu berücksichtigen, bei denen der Matrixeintrag für das letzte Phonem und das letzte Graphem nach Ausführen der Multiplikationen einen vorgegebenen Schwellwert überschreitet. Auf diese Weise werden lange Wörter ausgefiltert, bei denen die Zuordnung unsicher ist, ebenso wie sehr seltene und damit unsichere Zuordnungen.

[0025] Vorteilhafterweise wird für die Matrixeinträge eindeutiges Vorwissen genutzt, um stabile Fixpunkte zu schaffen. So wird z.B. der Matrixeintrag für das erste Phonem und das erste Graphem jedes Worts auf 1 gesetzt, ebenso wie der Matrixeintrag für das letzte Phonem und das letzte Graphem jedes Worts. Diese beiden Einträge bilden den Startpunkt bzw. Endpunkt des zu bestimmenden Weges und müssen auf jeden Fall durchlaufen werden. Andererseits werden der Matrixeintrag für das erste Phonem und das letzte Graphem jedes Worts sowie der Matrixeintrag für das letzte Phonem und das erste Graphem jedes Worts auf 0 gesetzt, denn diese Zuordnungen sind grundsätzlich ausgeschlossen.

[0026] Bei der Bestimmung des Maximums im Zusammenhang mit der Multiplikation wird die Diagonale als der wahrscheinlichste Weg bevorzugt. D.h., dass, falls bei der Bestimmung des Maximalwerts der drei vorhergehenden Matrixeinträge der Matrixeintrag für das vorhergehende Phonem und das vorhergehende Graphem im Wort und einer der anderen beiden Einträge gleich groß sind, der Matrixeintrag für das vorhergehende Phonem und das vorhergehende Graphem im Wort als maximal betrachtet wird.

[0027] Weitere vorteilhafte Weiterbildungen der Erfindung sind in den Unteransprüchen gekennzeichnet.

[0028] Im folgenden wird die Erfindung anhand von Ausführungsbeispielen näher erläutert, die in den Figuren schematisch dargestellt sind. Im einzelnen zeigt:

Fig. 1 ein zum Zuordnen von Phonemen zu den sie erzeugenden Graphemen in einem Lexikon geeignetes Computersystem;

Fig. 2 eine Matrix mit einer 1-zu-1 Zuordnung von Graphemen und Phonemen für das Wort <haben>;

Fig. 3 eine Matrix der Zuordnung von Graphemen und Phonemen für das Wort <textlich>;

Fig. 4 die Matrix der Übergangshäufigkeiten für die Zuordnung von Graphemen und Phonemen für das Wort <können>;

Fig. 5 die Matrix gemäß Fig. 4 nach Ausführen von Multiplikationen; und

Fig. 6A eine Matrix gemäß Fig. 5 für das Wort <yield>; und

Fig. 6B die Matrix gemäß Fig. 6A nach einer Korrektur der Zuordnung von Graphemen und Phonemen.

[0029] Fig. 1 zeigt ein zum Zuordnen von Phonemen zu den sie erzeugenden Graphemen geeignetes Computersystem. Dieses weist einen Prozessor (processor, CPU) 20, einen Arbeitsspeicher (RAM) 21, einen Programmspeicher (programm memory, ROM) 22, einen Festplatten-Controller (hard disc controller, HDC) 23, der eine Festplatte (hard disk) 30 steuert, und einen Schnittstellen-Controller (I/O controller) 24 auf. Prozessor 20, Arbeitsspeicher 21, Programmspeicher 22, Festplatten-Controller 23 und Schnittstellen-Controller 24 sind über einen Bus, den CPU-Bus 25, zum Austausch von Daten und Befehlen miteinander gekoppelt. Ferner weist der Computer einen Ein/Ausgabe-Bus (I/O Bus) 26 auf, der verschiedene Ein- und Ausgabeeinrichtungen mit dem Schnittstellen-Controller 24 koppelt. Zu den Ein- und Ausgabeeinrichtungen zählen z.B. eine allgemeine Ein- und Ausgabe-Schnittstelle (I/O interface) 27,

eine Anzeigeeinrichtung (display) 28, eine Tastatur (keyboard) 29 und eine Maus 31.

[0030] Im Folgenden wird beschrieben, wie die Zuordnung von Phonemen zu sie erzeugenden Graphemen für ein Wort durchgeführt wird.

[0031] In der folgenden Beschreibung werden verschiedene relative Häufigkeiten zur Berechnung der besten Zuordnung verwendet, die im Folgenden in der Regel kurz mit Häufigkeiten bezeichnet werden. Die Häufigkeit, mit der das Graphem g dem Phonem p zugeordnet wird, auch Übergangshäufigkeit genannt, berechnet sich aus

$$H(g \rightarrow p) = \frac{Z(g \rightarrow p)}{N(p)}$$

[0032] Dabei ist $Z(g \rightarrow p)$ die Anzahl der Zuordnungen des Graphems g , im Folgenden durch $\langle g \rangle$ bezeichnet, zu dem Phonem p , im Folgenden durch $[p]$ bezeichnet, und $N(p)$ die Anzahl aller Zuordnungen sämtlicher Grapheme zu diesem Phonem $[p]$.

[0033] Es werden noch weitere Häufigkeiten benötigt, da zur endgültigen Entscheidung über die Zuordnungen die relative Häufigkeit der direkten Zuordnung eines Graphems zu einem Phonem nicht ausreicht. Daher werden noch positionsabhängige Häufigkeiten in Graphemgruppen $\langle G \rangle$ sowie die Vorgänger- und Nachfolgerhäufigkeiten bestimmt, die die Abhängigkeiten der Zuordnung zu Phonemen von den vorhergehenden und nachfolgenden Graphemen widerspiegeln.

[0034] Unter einer positionsabhängigen Häufigkeit H^{pos} versteht man die Häufigkeit, mit der das Graphem an einer bestimmten Position innerhalb einer Graphemgruppe $\langle G \rangle$ einem Phonem zugeordnet wird. So findet sich z.B. bei der Zuordnung der Graphemgruppe $\langle ch \rangle$ zum Phonem $[C]$ das Graphem $\langle c \rangle$ an erster und das Graphem $\langle h \rangle$ an zweiter Position. Dabei ist $[C]$ der stimmlose palatale Frikativ oder Ich-Laut, wie in $\langle \text{Sicht} \rangle$.

[0035] Die Häufigkeit H^{pos} berechnet sich aus

$$H^{pos}(g \rightarrow p | g \text{ in } \langle G \rangle \text{ an Pos } i) = \frac{Z(g \rightarrow p | g \text{ in } \langle G \rangle \text{ an Pos } i)}{N(p)}$$

[0036] Zur Initialisierung der Übergangshäufigkeiten werden die Einträge in einem Lexikon mit Wörtern und deren Lautschrift verwendet, bei denen die Anzahl der Grapheme mit der Anzahl der Phoneme übereinstimmt. Es wird angenommen, dass jedes Graphem dem entsprechenden Phonem zugeordnet ist. Dies ist in Fig. 2 durch die diagonal verlaufende Linie dargestellt.

[0037] Diese direkte Zuordnung ist nicht immer korrekt, wie etwa das Beispiel $\langle \text{textlich} \rangle$ aus Fig. 3 zeigt, in der die Linie für die Zuordnungen nicht einfach diagonal verläuft. Die Anzahl der Grapheme im Wort $\langle \text{textlich} \rangle$ stimmt mit der Anzahl der Phoneme überein. Es sind jeweils 8. Jedoch wird der Buchstabe $\langle x \rangle$ auf zwei Phoneme $[ks]$ abgebildet und die Buchstabengruppe $\langle ch \rangle$ auf nur ein Phonem $[C]$. Da derartige Ausnahmen jedoch relativ selten auftreten, werden sie bei der Anwendung der relativen Häufigkeiten entsprechend gering gewichtet. Außerdem werden alle Häufigkeiten, die einen bestimmten Schwellwert unterschreiten, in einem späteren Korrekturschritt entfernt.

[0038] Die Zuordnungen werden gezählt, und aus ihnen werden die relativen Häufigkeiten bzw. Übergangshäufigkeiten bestimmt.

[0039] Mit den im vorhergehenden Schritt gewonnenen relativen Häufigkeiten oder Übergangshäufigkeiten wird für jedes Wort im Lexikon eine Matrix mit Übergangshäufigkeiten aufgestellt, wie sie in Fig. 4 für das Wort $\langle \text{können} \rangle$ gezeigt ist.

[0040] Dabei werden vier Einträge fest vorgegeben. Die Einträge links unten und rechts oben müssen immer durchlaufen werden, da sie Start- bzw. Endpunkt sind. Sie werden deswegen auf 1 gesetzt. Dagegen können die Felder links oben und rechts unten niemals durchlaufen werden. Sie werden daher auf 0 gesetzt. Alle anderen Felder enthalten die entsprechenden Übergangshäufigkeiten $H(g \rightarrow p)$.

[0041] Bei dieser initialen Zuordnung wurde $\langle n \rangle$ dem Phonem $[ŋ]$ (gerundeter halboffener vorderer Vokal "ö") zugeordnet. Daher steht in den entsprechenden Feldern keine 0, sondern 0.013. Man sieht aber, dass diese Häufigkeit viel kleiner ist als die übrigen Häufigkeiten. Sie fällt demnach kaum ins Gewicht.

[0042] Für die Berechnung des Wegs werden nun die einzelnen Matrixeinträge jeweils mit dem Maximum der Nachbareinträge multipliziert. Da nur die Bewegungen nach oben, nach rechts oder nach oben rechts erlaubt sind, werden nur die Werte links, unten und links unten vom jeweiligen Matrixeintrag aus für die Bestimmung des Maximums betrachtet.

[0043] Falls bei der Bestimmung des Maximalwerts der Matrixeintrag links unten (diagonal) vom jeweiligen Matrixeintrag aus und einer der anderen beiden Einträge gleich groß sind, wird der diagonal liegende Matrixeintrag als maximal betrachtet.

[0044] Bei der Multiplikation wird mit dem ersten Eintrag unten links begonnen, wobei bei der Bestimmung der Maximalwerte die sich jeweils aus den Multiplikationen ergebenden geänderten Einträge der Matrix herangezogen werden.

[0045] Die erste Spalte und die unterste Zeile stellen Sonderfälle dar, da es keinen linken bzw. unteren Nachbarn gibt. Hier wird der aktuelle Eintrag stets mit dem unteren bzw. dem linken Eintrag multipliziert. Die einzelnen sich ergebenden Produkte sind in Fig. 5 dargestellt.

[0046] Die akkumulierte Häufigkeit am Endpunkt oben rechts ist somit das Produkt der Einträge bzw. Häufigkeiten auf dem optimalen Weg vom Startpunkt zum Endpunkt.

[0047] Durch das Bestimmen, welcher der drei vorhergehenden Matrixeinträge maximal war, wird eine Schrittrichtung von Matrixeintrag zu Matrixeintrag ermittelt. Ausgehend vom Matrixeintrag für das letzte Phonem und das letzte Graphem (oben rechts) wird jeweils entlang der ermittelten Schrittrichtung ein Weg durch die Matrix definiert bis zum Matrixeintrag unten links. Die zu dem Weg gehörenden Matrixelemente definieren die Zuordnung von Graphemen zu Phonemen des Worts.

[0048] Anschließend erfolgt eine Nachbehandlung zu weiteren Verbesserung der Zuordnung. Die Nachbehandlung dient zum Überprüfen der getroffenen Entscheidungen unter Berücksichtigung des Graphem- und Phonem-Kontextes.

[0049] Zunächst wird nach Ausführen der beschriebenen Zuordnung von Graphemen zu Phonemen für jedes Wort des Lexikons aus diesen Zuordnungen die relative Häufigkeit ermittelt, mit der ein Phonem durch zwei oder mehr Grapheme erzeugt wird bzw. zwei oder mehr Phoneme durch ein Graphem erzeugt werden, d.h. die positionsabhängige Häufigkeit H^{pos} .

[0050] Anschließend wird die Zuordnung von Graphemen zu Phonemen innerhalb eines Worts mit Hilfe der positionsabhängigen Häufigkeiten korrigiert. Dazu betrachte man Fig. 6A, die im Aufbau Fig. 5 entspricht. Das bisher beschriebene Verfahren liefert z.B. für das englische Wort <yield> die Zuordnung

	y	e	l	d
zu	j	i:	l	d

da die Häufigkeit der Zuordnung des Graphems <i> zum Phonem [j] höher ist (hier 0,04) als die Häufigkeit der Zuordnung zum Phonem [i:] (hier 0,03).

[0051] Die positionsabhängigen Häufigkeiten zeigen jedoch, daß die Häufigkeit der Zuordnung von <i> zum Phonem [j] klein ist, wenn <i> sich an zweiter Position der Graphemgruppe <yi> befindet. Hingegen ist die Häufigkeit der Zuordnung von <i> zum Phonem [i:] groß, wenn <i> sich an erster Position der Graphemgruppe <ie> befindet.

[0052] Unterstützt wird diese korrigierte Zuordnung auch durch die Betrachtung der positionsabhängigen Häufigkeit von <e>. Die Häufigkeit der Zuordnung von <e> zum Phonem [i:] ist klein, wenn <e> sich vor <l> befindet. Hingegen ist die Häufigkeit der Zuordnung von <e> zum Phonem [i:] groß, wenn <e> sich an zweiter Position der Graphemgruppe <ie> befindet.

[0053] Entsprechend Fig. 6B kann somit die Zuordnung korrigiert werden.

[0054] Nach Ausführen der korrigierten Zuordnung für jedes Wort des Lexikons werden aus diesen korrigierten Zuordnungen die Übergangshäufigkeiten und die positionsabhängigen Häufigkeiten ermittelt. Diese werden bei weiteren Zuordnungen verwendet.

[0055] Zum Bestimmen der relativen Häufigkeiten werden nur solche Zuordnungen berücksichtigt, bei denen der Matrixeintrag für das letzte Phonem und das letzte Graphem (oben rechts) nach Ausführen der geschilderten Multiplikationen einen vorgegebenen Schwellwert überschreitet. Dieser Matrixeintrag entspricht dem Produkt der Übergangshäufigkeiten entlang des besten Wegs. Die Größe dieses Produkts wird somit als Kriterium verwendet, ob dieser Weg akzeptiert werden soll oder nicht.

[0056] Das Verfahren wird in mehreren Iterationen ausgeführt. Dabei ist der Schwellwert am Anfang hoch und wird nach jeder Iteration verkleinert. Dadurch werden am Anfang nur solche Zuordnungen akzeptiert, die relativ sicher richtig sind. Da alle Häufigkeiten kleiner gleich 1 sind, fließt indirekt in das Produkt auch die Länge des Wortes ein. Je mehr Faktoren das Produkt hat, desto kleiner wird es. Somit werden anfangs vorwiegend die Zuordnungen von kurzen Wörtern akzeptiert. Bei kurzen Wörtern ist die Wahrscheinlichkeit, eine falsche Zuordnung zu finden, kleiner als bei langen.

[0057] Die Zuordnungen, bei denen das Produkt der Übergangshäufigkeiten den Schwellwert überschritten hat, werden zur Gewinnung der neuen Statistik herangezogen. Schon bei der ersten Auswertung der so gewonnenen Statistik sind die meisten Fehler verschwunden, die sich durch die eins zu eins Initialisierung der Häufigkeiten ergaben. Außerdem wird noch überprüft, wie häufig jede Graphem-Phonem-Zuordnung auftrat. Wenn das Verhältnis einen Schwellwert unterschreitet, wird diese Zuordnung ignoriert und somit beim nächsten Auffüllen der Matrizen nicht weiter verwendet.

[0058] Im Ergebnis erhält man eine Zuordnung der Grapheme zu den Phonemen für das gesamte Lexikon. Ferner erhält man eine Aufstellung, welches Phonem bzw. welche Phonemgruppe durch welche Grapheme erzeugt werden kann, z.B. [tʃ] im Englischen durch <ch>, <cz>, <c>, <tch>, <cc>, <t> und <che>.

Patentansprüche

1. Verfahren zum Zuordnen von Phonemen zu den sie erzeugenden Graphemen in einem Lexikon mit Wörtern (Graphemfolgen) und deren zugehöriger Lautschrift (Phonemfolgen),
 5 **dadurch gekennzeichnet,**
dass die Zuordnung mit Hilfe eines Algorithmus der dynamischen Zeitanpassung (DTW, Dynamic Time Warping) durchgeführt wird.
2. Verfahren zum Zuordnen von Phonemen zu ihnen zuzuordnenden Graphemen in einem Lexikon mit Wörtern (Graphemfolgen) und deren zugehöriger Lautschrift (Phonemfolgen):
 10
- a) In einem ersten Schritt wird für jede Zuordnung von Phonemen und Graphemen eine relative Häufigkeit ermittelt, mit der Phoneme und Grapheme einander zugeordnet werden,
 - b) in einem zweiten Schritt wird für jedes Wort des Lexikons eine zweidimensionale Matrix (Inzidenzmatrix) erstellt, deren einer Index durch die Grapheme des Worts und deren zweiter Index durch die Phoneme des Worts gegeben ist,
 - c) als Einträge der Matrix werden die zu dem jeweiligen Phonem-Graphem-Paar gehörenden, im ersten Schritt bestimmten, relativen Häufigkeiten gewählt,
 - d) in einem dritten Schritt wird jeder Matrixeintrag mit einer mathematischen Operation mit dem Extremalwert der folgenden drei vorhergehenden Matrixeinträge verknüpft:
 20
 - dem Eintrag für dasselbe Phonem und das vorhergehende Graphem im Wort,
 - dem Eintrag für das vorhergehende Phonem und dasselbe Graphem im Wort und
 - dem Eintrag für das vorhergehende Phonem und das vorhergehende Graphem im Wort,
 - e) bei der mathematischen Operation wird mit dem ersten Graphem und dem ersten Phonem des Worts begonnen, wobei bei der Bestimmung der Extremalwerte die sich jeweils aus den mathematischen Operationen ergebenden geänderten Einträge der Matrix herangezogen werden,
 - f) durch das Bestimmen, welcher der drei vorhergehenden Matrixeinträge eines Matrixeintrags extremal war, wird eine Schrittrichtung für diesen Matrixeintrag ermittelt,
 - g) in einem vierten Schritt wird ausgehend vom Matrixeintrag für das letzte Phonem und das letzte Graphem jeweils die für den Matrixeintrag ermittelte Schrittrichtung entlang ein Weg durch die Matrix definiert bis zum Matrixeintrag für das erste Phonem und das erste Graphem,
 - h) die zu dem Weg gehörenden Matrixelemente definieren die Zuordnung von Graphemen zu Phonemen des Worts.
 35
3. Verfahren nach mindestens einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
dass die relativen Häufigkeiten im ersten Schritt dadurch ermittelt werden, dass Wörter aus dem Lexikon gewählt werden, bei denen die Anzahl der Grapheme und die Anzahl der Phoneme übereinstimmen, und dass bei diesen Wörtern die Grapheme und Phoneme in der Reihenfolge der Angabe ihrer Grapheme und Phoneme im Lexikon einander zugeordnet werden.
 40
4. Verfahren nach mindestens einem der vorhergehenden Ansprüche,
 45 **dadurch gekennzeichnet,**
dass, nach Ausführen der Zuordnung von Graphemen zu Phonemen gemäß dem Verfahren nach Anspruch 1 für jedes Wort des Lexikons, aus diesen Zuordnungen die positionsabhängige relative Häufigkeit ermittelt wird, mit der ein Phonem durch zwei oder mehr Grapheme erzeugt wird oder zwei oder mehr Phoneme durch ein Graphem erzeugt werden oder zwei oder mehr Grapheme einem Phonem zugeordnet werden oder ein Graphem zwei oder mehr Phonemen zugeordnet wird.
 50
5. Verfahren nach Anspruch 4,
dadurch gekennzeichnet,
dass die Zuordnung von Graphemen zu Phonemen innerhalb eines Worts gemäß Anspruch 1 mit Hilfe der positionsabhängigen relativen Häufigkeiten korrigiert werden.
 55
6. Verfahren nach Anspruch 5,
dadurch gekennzeichnet,

dass, nach Ausführen der Zuordnung von Graphemen zu Phonemen gemäß dem Verfahren nach Anspruch 3 für jedes Wort des Lexikons, aus diesen korrigierten Zuordnungen die positionsabhängige relative Häufigkeit ermittelt wird, mit der ein Phonem durch zwei oder mehr Grapheme erzeugt wird bzw. zwei oder mehr Phoneme durch ein Graphem erzeugt werden; und

dass in Anspruch 3 anschließend die neu berechneten relativen Häufigkeiten verwendet werden.

7. Verfahren nach Anspruch 6,

dadurch gekennzeichnet,

dass zum Bestimmen der relativen Häufigkeiten nur solche Zuordnungen berücksichtigt werden, bei denen der Matrixeintrag für das letzte Phonem und das letzte Graphem nach Ausführen der Multiplikationen einen vorgegebenen Schwellwert überschreitet.

8. Verfahren nach mindestens einem der vorhergehenden Ansprüche,

dadurch gekennzeichnet,

dass der Matrixeintrag für das erste Phonem und das erste Graphem jedes Worts auf 1 gesetzt wird;

dass der Matrixeintrag für das letzte Phonem und das letzte Graphem jedes Worts auf 1 gesetzt wird;

dass der Matrixeintrag für das erste Phonem und das letzte Graphem jedes Worts auf 0 gesetzt wird; und

dass der Matrixeintrag für das letzte Phonem und das erste Graphem jedes Worts auf 0 gesetzt wird.

9. Verfahren nach mindestens einem der vorhergehenden Ansprüche,

dadurch gekennzeichnet,

dass, falls bei der Bestimmung des Maximalwerts der drei vorhergehenden Matrixeinträge der Matrixeintrag für das vorhergehende Phonem und das vorhergehende Graphem im Wort und einer der anderen beiden Einträge gleich groß sind, der Matrixeintrag für das vorhergehende Phonem und das vorhergehende Graphem im Wort als maximal betrachtet wird.

10. Computerprogrammprodukt, das durch einen Computer ausführbar ist und dabei die Schritte nach einem der Ansprüche 1 bis 9 ausführt.

11. Computerprogrammprodukt, das auf einem computergereinigtem Medium gespeichert ist und computerlesbare Programmmittel umfasst, die es einem Computer ermöglichen, das Verfahren nach einem der Ansprüche 1 bis 9 auszuführen.

12. Datenträger, auf dem ein Computerprogramm gespeichert ist, das es einem Computer ermöglicht, durch einen Ladeprozess das Verfahren nach einem der Ansprüche 1 bis 9 auszuführen.

13. Computersystem zum Zuordnen von Phonemen zu den sie erzeugenden Graphemen in einem Lexikon mit Wörtern (Graphemfolgen) und deren zugehöriger Lautschrift (Phonemfolgen)

- mit einer Speichereinrichtung (22, 30) zum Speichern eines Computerprogramms auf einem Speichermedium;
- mit einer Verarbeitungseinheit (20) zum Laden des Computerprogramms aus der Speichereinrichtung und zum Ausführen des Computerprogramms;
- mit Mitteln zum Durchführen der Zuordnung mit Hilfe eines Algorithmus der dynamischen Zeitanpassung (DTW, Dynamic Time Warping).

FIG 1

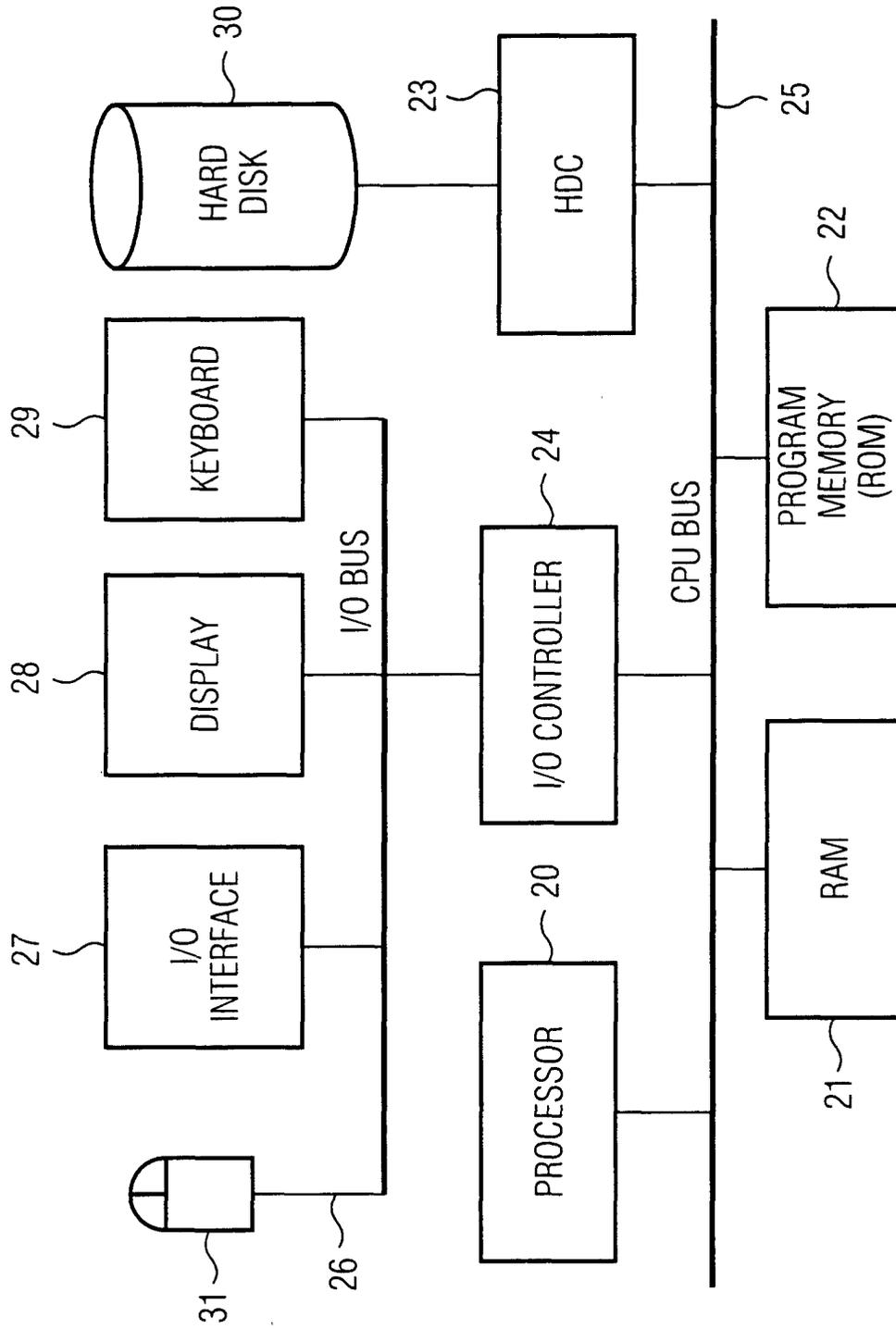


FIG 2

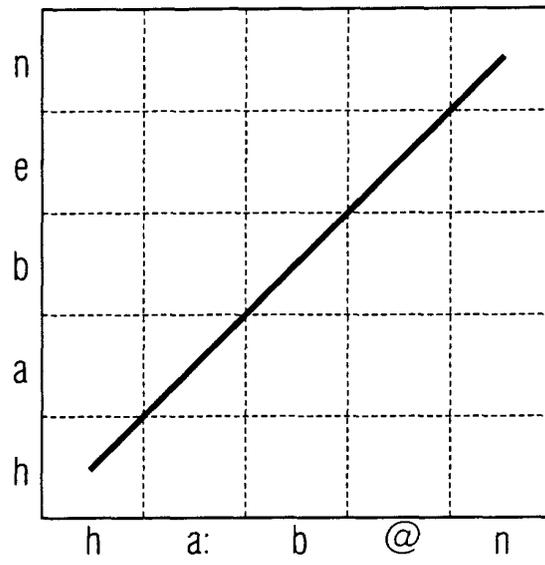


FIG 3

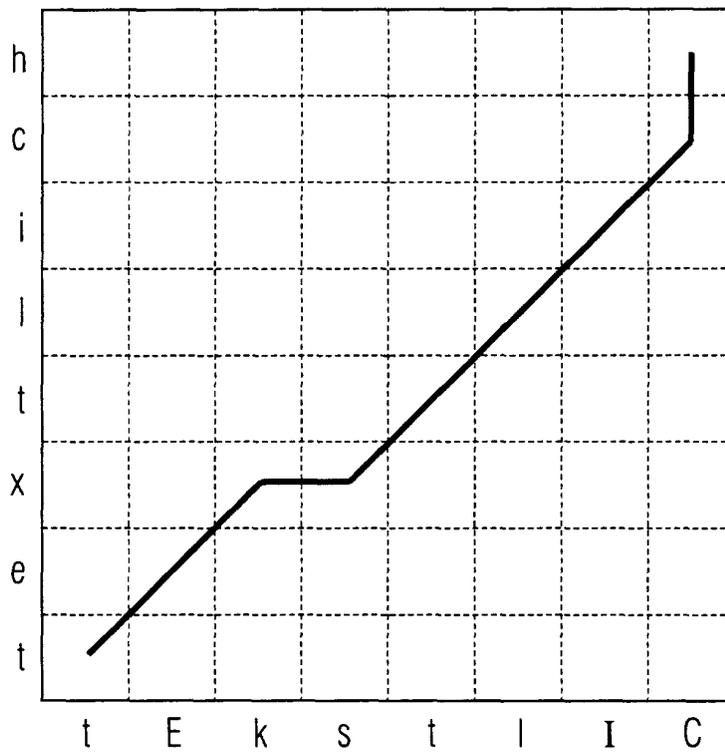


FIG 4

n	0	0.013	0.949	0	1
e	0	0	0	0.986	0
n	0	0.013	0.949	0	0.949
n	0	0.013	0.949	0	0.949
ö	0	0.920	0	0	0
k	1	0	0	0	0
	k	9	n	@	n

FIG 5

n	0	0	0	0	0.817
e	0	0	0	0.817	0
n	0	0.000	0.828	0	0
n	0	0.012	0.873	0	0
ö	0	0.920	0	0	0
k	1	0	0	0	0
	k	9	n	@	n

FIG 6A

d	0	0	0	0.01
l	0	0	0.01	0
e	0.00	0.01	0.00	0
i	0.04	0.03	0	0
y	1.0	0.00	0	0
	j	i:	l	d

FIG 6B

d	0	0	0	0.01
l	0	0	0.01	0
e	0.00	0.01	0.00	0
i	0.04	0.03	0	0
y	1.0	0.00	0	0
	j	i:	l	d