

VOKABULARREPRÄSENTATION ZUR AUTOMATISCHEN IDENTIFIKATION BEI DAWID II

KURT KOTTEN

Die automatische Identifizierung gesprochener Wörter mithilfe des Systems DAWID II (Device for Automatic Word Identification by Discrimination) geschieht in zwei aufeinanderfolgenden Schritten, die von einem hardware-Analysator und einem software-Identifikator oder Zuordner durchgeführt werden. Der Analysator segmentiert das kontinuierliche akustische Signal des gesprochenen Wortes. Er kann z.Zt. zehn Segmentklassen unterscheiden, die wir Auteme (Tillman 1967) (Automatenphoneme) nennen. Da es sich bei den Autemen nicht etwa um eine Auswahl normaler Phoneme handelt, sind sie nicht durch phonetische Symbole, sondern einfach durch die Zahlen 1 bis 10 symbolisiert. Der Wert dieser Zahlen spielt keine Rolle, es handelt sich um willkürlich zugeordnete Symbole.

Jede Realisation eines Wortes erzeugt am Ausgang des Analysators eine solche Symbolfolge, eben eine Autemfolge. Verschiedene Realisationen ein und desselben Wortes erzeugen (möglicherweise) verschiedene, aber einander ähnliche Autemfolgen. Der Zuordner hat dann zunächst die Aufgabe, Gemeinsamkeiten aller vorliegenden Realisationen eines Wortes, d.h. aller zu diesem Wort gehörenden Autemfolgen zu finden. Als Ergebnis erhält man eine Normfolge (Kotten 1971) für dieses Wort, die zusammen mit dem Klartext ins Vokabular abgespeichert wird. Nachdem alle Wörter, die zu einem beliebig festgelegten Vokabular gehören, auf diese Weise bearbeitet wurden, kann der Benutzer zum eigentlichen Identifikationsprozeß übergehen. Jetzt wird jede unbekannte Autemfolge mit allen im Vokabular gespeicherten Normfolgen verglichen und als Output erscheint dasjenige Wort im Klartext, zu dem die ähnlichste Normfolge gehört. Aufgrund dieser Arbeitsweise heißt das Programmpaket NOFI (Normfolgenidentifikator). Es handelt sich um ein Mensch-Machine-Interaktionssystem, das auf der PDP 15/20 unseres Institutes programmiert wurde. Zur Unterscheidung der zwei Phasen der Verarbeitung bei NOFI sprechen wir wie üblich von Lern- und Kannphasen. In beiden Phasen brauchen wir ein Maß zum Vergleich von Symbolketten. Zum gegenwärtigen Zeitpunkt scheint es auszureichen, nicht ein symmetrisches Ähnlichkeitsmaß zu definieren, sondern nur danach zu fragen, ob eine Symbolkette in einer anderen enthalten ist und in welchem Maße. Wir definieren deshalb als Vergleichsmaß zweier Ketten (a) und (b)

$$P = P_1 \cdot P_2$$

$$P_1 = \text{Länge (a)} / \text{Länge (b)}; 0 < P_1 \leq 1$$

$$P_2 = \text{Länge (a)} / A \quad ; \quad 0 < P_2 \leq 1$$

A = Abstand des ersten vom letzten Element aus (a) in (b)

Insbesondere in der Definition von P_1 kommt die Asymmetrie des Maßes zum Ausdruck. Es wird überhaupt nur dann verglichen, wenn Länge (a) \leq Länge (b) ist. Hierin kommt die Hypothese zum Ausdruck, daß Normfolgen, die ja die Gemeinsamkeiten aller Autemfolgen eines Wortes repräsentieren sollen, kürzer sind als die Autemfolge einer konkreten Realisation. Ebenso wie für P_1 gilt auch für P_2 eine Einschränkung. P_2 wird nur berechnet, wenn alle Elemente von a in derselben Reihenfolge in (b) vorkommen.

Ein weiteres Charakteristikum des P -Maßes ist, daß identische Symbolketten den Wert $P = 1$ erzeugen. Am folgenden Beispiel soll die Berechnung von P demonstriert werden.

(a): 7 3 10 7 ; Länge (a) = 4
 1 \ \ \
 7 10 3 10 7 10; Länge (b) = 6 ; $A = 5$

$$P_1 = 4/6 ; P_2 = 4/5 ; P = (4/6) \cdot (4/5) = 0.533..$$

Am Beispiel des Wortes 'fünf' soll jetzt beschrieben werden, wie in der Lernphase eine Normfolge berechnet wird. Gegeben sind folgende elf Autemfolgen dieses Wortes.

- (1) 6 10 3 10 6
- (2) 7 10 3 10 7
- (3) 6 1 3 9 3 1 6
- (4) 7 10 3
- (5) 7 10 3 10 7
- (6) 7 3 10 7
- (7) 7 3 7
- (8) 7 3 7
- (9) 7 10 3 10 7
- (10) 7 3 7
- (11) 7 3 10 7

Jede dieser Folgen wird mit jeder anderen verglichen und die Ergebnisse in Form einer P -Matrix aufgeschrieben. Ist i der Zeilenindex und j der Spaltenindex, so bedeutet P_{ij} , daß die i -te Folge mit der j -ten Folge verglichen wurde. Ist eine Folge in keiner anderen enthalten, so entsteht eine Matrixzeile mit lauter Nullen. Ist dagegen eine Folge so beschaffen, daß keine andere in ihr enthalten ist, so entsteht eine Spalte mit lauter Nullen. In unserem Beispiel sind das die Spalten 1,3 und 4. Im nächsten

Schritt werden sowohl Zeilen- als auch Spaltensummen dieser Matrix berechnet. Ist eine der Spaltensummen gleich Null, so wird die Folge, die zu dieser Spalte gehört, eliminiert. Dann werden in den verbleibenden Folgen diejenigen Auteme gestrichen, die nicht in all diesen Folgen vorkommen. Im Beispiel ist nach der Streichung der Folgen 1,3 und 4 das Autem 10 nicht in allen verbleibenden Folgen enthalten. Es wird deshalb vor der Berechnung einer neuen P -Matrix eliminiert. In der neuen P -Matrix ist keine der Spaltensummen gleich Null. Erst wenn dieser Fall eintritt, wird zur Bestimmung der endgültigen Normfolge das Maximum der Zeilensummen gesucht. Auf diese Weise ist diejenige (möglicherweise gekürzte) Folge gefunden, die den meisten anderen am ähnlichsten ist. In unserem Beispiel heißt die Normfolge '7 3 7'. Der Vorteil dieser Methode besteht darin, daß bei genügend großem Lernmaterial Irrtümer (Versprecher des Benutzers oder Irregularitäten des Analysators) automatisch eliminiert werden können, so daß das System auch für ungeübte Benutzer hinreichend sicher zu arbeiten imstande ist.

Da jeder Benutzer sein Vokabular frei wählen kann, kann es vorkommen, daß zwei Wörter dieselbe Normfolge haben. Deshalb ist bei NOFI zwischen die Lernphase und die Kannphase eine sog. Editionsphase eingeschoben, in der das Normfolgevokabular auf solche Fälle hin untersucht werden kann. Haben z.B. die Wörter 'null' und 'neun' dieselbe Normfolge, so muß eins dieser Wörter aus dem Vokabular gestrichen werden. Man kann z.B. das Wort 'null' streichen und statt dessen, indem man wieder zur Lernphase zurückgeht, das Wort 'zero' an seiner Stelle ins Vokabular aufnehmen. Da es sich um ein Interaktionssystem handelt, ist ein Übergang von einer Phase in die andere jederzeit möglich.

DAWID I bestand aus einem hardware-Analysator und einem hardware-Zuordner. Er konnte ein festes Vokabular von zehn Zahlwörtern und zehn Rechenmaschinenebefehlen sprecherunabhängig identifizieren. Bei DAWID II ist durch den hardware-Analysator nach wie vor die Sprecherunabhängigkeit gegeben, durch den software-Identifikator wird nun zusätzlich eine Vokabularindifferenz erreicht, die es jedem Benutzer freistellt, sein eigenes Vokabular aufzubauen. Dieses wird unter seinem Namen auf Magnetband abgespeichert.

*Institut für Kommunikationsforschung und Phonetik
Bonn*

LITERATUR

- Kotten, K.
1967 "Nicht-adaptive und adaptive Sprachsignal-Identifikation", *IPK-Forschungsbericht* 68.1 (Hamburg, H. Buske Verlag).
- 1971 "Zwei Konzeptionen der Zuordnungslogik von DAWID II", in *DAWID II-Beiträge zur automatischen Spracherkennung, IPK-Forschungsberichte* 39 (Hamburg, H. Buske Verlag).
- Tillmann, H.G.
1967 "Akustische Phonetik und linguistische Akustik", *Phonetica* 16:143-155.

DISCUSSION

IIVONEN (Oulu)

Sie haben eigentlich nichts über die Beschaffenheit der Auteme gesagt. Die Theorie der Auteme scheint wichtig auch für die Theorie der akustischen Phonetik zu sein. Was würden Sie über den Autembestand sagen, ist er schon fertig oder wird er vielleicht noch weiter entwickelt?

KOTTEN

Die Definition des Autem-Begriffs wurde von Tillmann in der angegebenen Literatur dargelegt. Auf die Beschaffenheit der Auteme von DAWID II einzugehen war in der kurzen Zeit des Vortrages nicht möglich. Ich möchte hier nur auf den Vortrag von Rupprath (to be found on pp. 000-000 of this volume) über das [a,ɔ] — Autem hinweisen. Es wäre vermessen zu sagen, dass der Autembestand schon vollständig sei. Mein System NOFI versucht, mit dem gegebenen Bestand von Autemen eine möglichst grosse Zahl von Wörtern zu erkennen. Es braucht jedoch auch bei beliebiger Erweiterung des Autembestandes nicht geändert werden. Ein konkretes Beispiel für eine mögliche Erweiterung wurde von Vieregge/Glave in *Section B1 dargestellt*.