

Lutz Friedemann

Technische Universität Dresden

## 1. Einleitung

In der Arbeitsgruppe Kommunikation der TU Dresden haben wir uns in den letzten Jahren zunehmend damit beschäftigt, die Spracherkennung nicht ausschließlich mit formalen Algorithmen durchzuführen. Es hat sich gezeigt, daß selbst bei Optimierung der Vorverarbeitung eine bestimmte Grenze der Erkennungsrate nicht zu überschreiten ist. Dabei ist diese Optimierung sprecher- und wortschatzabhängig durchzuführen. Ausgehend vom natürlichen Vorbild - dem Menschen - wurde nun versucht, Erkennungsalgorithmen in Form von Expertensystemen /1/ zu entwickeln, die über ein psychoakustisches Hintergrundwissen verfügen. Bei den Untersuchungen stellte sich heraus, daß unscharfe Beschreibungsformen durchaus günstig in allen Ebenen eines solchen Expertensystems angewendet werden können. Diese Beschreibungsformen liefern die Möglichkeit, eine Entscheidung nur unscharf aber mit einer bestimmten Wichtung zu treffen. Außerdem liefern sie auch Verknüpfungsvorschriften, die eine Weiterverarbeitung dieser unscharfen Entscheidung ermöglichen. Mit dieser Methode besitzt man ähnliche Möglichkeiten, Entscheidungen zu treffen, wie bei der Wahrscheinlichkeitsrechnung, jedoch ist die Forderung aufgehoben, daß die Summe aller Zugehörigkeiten gleich "1" sein muß. Da die Einführung der oben genannten Beschreibung im Expertensystem EXA 1630 /2/ zu ermutigenden Ergebnissen führte, wurde versucht, diese Methode auf die Vorverarbeitung in einem formalen Erkenneralgorithmus anzuwenden. Die Zielstellung war, eine Vektorquantisierung zu erhalten, die sich dadurch auszeichnet, daß am Eingang anliegende spektrale Beschreibungen mittels bereits bekannter Filteralgorithmen und unscharfer Beschreibung in Symbolzugehörigkeitsfolgen umgewandelt werden. Diese Quantisierung soll zunächst im Lern- und Erkennungsmodus erfolgen.

## 2. Abstandsmaß und Zugehörigkeit

Alle Erklärungen erfolgen im 2-dimensionalen Merkmalraum. Dies ist keine Beschränkung der Allgemeinheit. Es dient nur dem besseren Verständnis.

Die verwendeten Abstandsmaße - City-Block-Distanz, euklidischer Abstand - zeichnen sich dadurch aus, daß sie im mathematischen Sinne Maße sind. Sie sind z.B. symmetrisch, d.h. der Abstand eines Punktes A vom Punkt B ist gleich dem Abstand des Punktes B vom Punkt A. Diese sehr nützliche Eigenschaft widerspricht aber in der Spracherkennung der Tatsache, daß Verwechslungen beim Menschen häufiger unsymmetrisch sind. Betrachtet man die Linien gleichen Abstandes von einem Punkt A im 2-dimensionalen Merkmalraum, so ergeben sich bei der City-Block-Distanz Geradenstücken, die Teil eines Quadrates sind. Daraus folgt, daß nur Merkmalballungen gut beschrieben werden können, die eine solche Form besitzen. Führt man 2 Klassen A und B ein, die auf der Basis dieses Abstandes getrennt werden sollen, so ergibt sich als Trennfläche zwischen diesen Klassen eine Gerade. Eine solche Trenngerade ergibt sich bei allen Abstandsmaßen im 2-dimensionalen Merkmalraum, da diese symmetrisch sind.

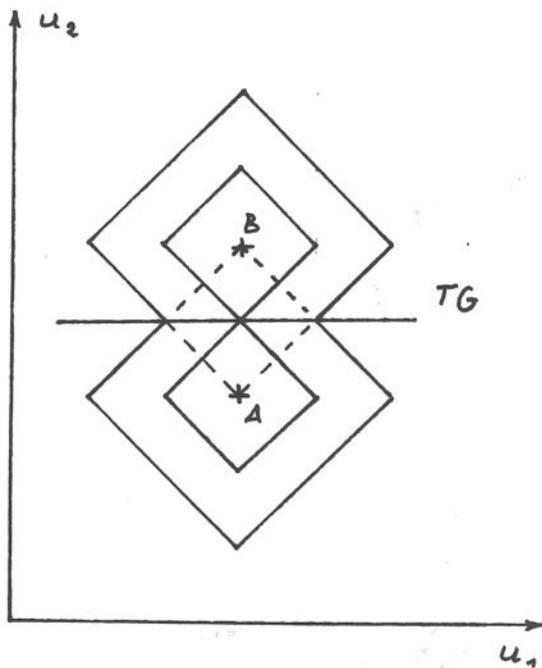


Bild 1: Linien gleichen Abstandes und Trenngerade

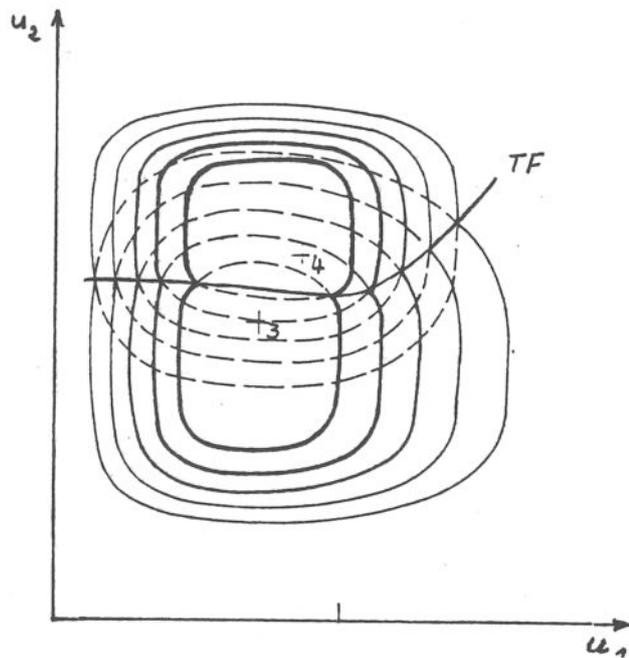
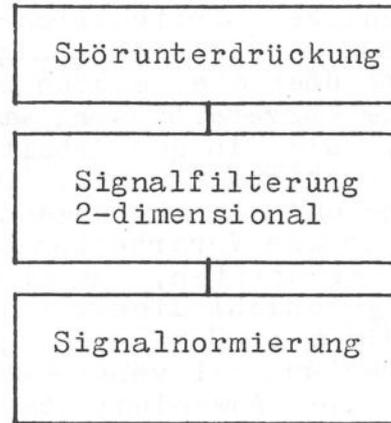
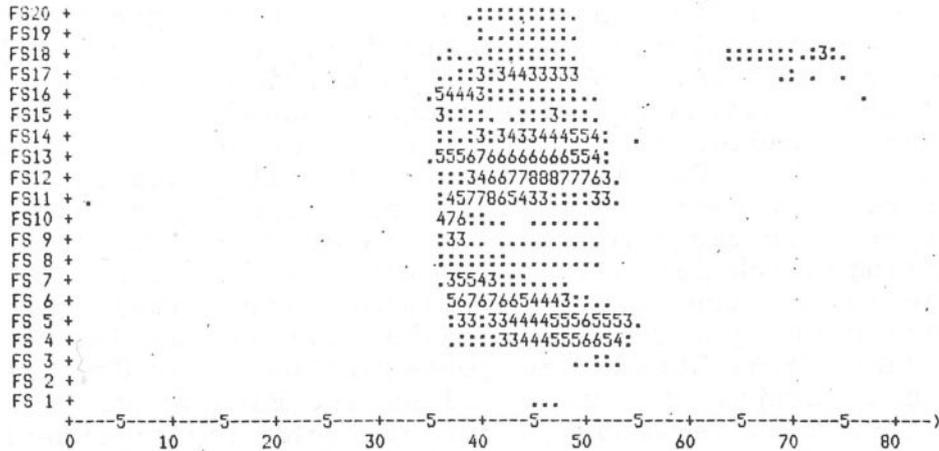


Bild 2: Linien gleicher Ähnlichkeit und Trennfläche

Führt man nun die unscharfe Beschreibung ein, werden in unserem Fall Ähnlichkeiten betrachtet. Dies hat für die Entscheidung in der Objekterkennung keine Konsequenzen, da sich Entscheidungen auf der Basis der Unähnlichkeit in Entscheidungen auf der Basis der Ähnlichkeit überführen lassen /3/. Betrachtet man also die Linien gleicher Ähnlichkeit im 2-dimensionalen Merkmalraum, so ergeben sich konvexe Flächen. Diese müssen aufgrund der gewählten Beschreibung nicht notwendig symmetrisch sein. Für die Beschreibung wurden die von Boklich vorgeschlagenen, modifizierten Potentialfunktionen verwendet /4/. Diese Beschreibung hat den Vorteil, daß eine Vielzahl konvexer Ballungen im Merkmalraum modelliert werden kann. Das Konzept bietet auch die Möglichkeit, über eine Verknüpfung mehrerer konvexer Teilbeschreibungen zu einer konkaven Gesamtbeschreibung einer Klasse zu gelangen. Im Bild 2 erkennt man deutlich, daß die Ähnlichkeit des Punktes A bezüglich der Klasse B sehr stark von der Ähnlichkeit des Punktes B bezüglich der Klasse A abweicht. Diese Unsymmetrie bewirkt, daß es sich bei der Ähnlichkeit nicht mehr um Maße handelt. Damit muß jede weitere Operation mit diesen Zahlen erst auf ihre allgemeine Gültigkeit hin untersucht werden. Betrachtet man nun im 2-dimensionalen Merkmalraum zwei Klassen, die so beschrieben sind, und legt eine Trennfläche zwischen diese beiden Klassen, so ergibt sich im Allgemeinen eine krummlinige Kurve. Aufgrund der genannten Eigenschaften erschien diese Beschreibungsform besser geeignet für einen formalen Erkennungsalgorithmus als die bisherige Abstandsbeschreibung.

SUMM 12222222222222222222 1 11 11  
 5 15555555555555555555182 6 2698070373087142  
 281 3333333333333333333364 22 5457927820601843



DMO:W1S3T1.N02;1 001:001  
EINS

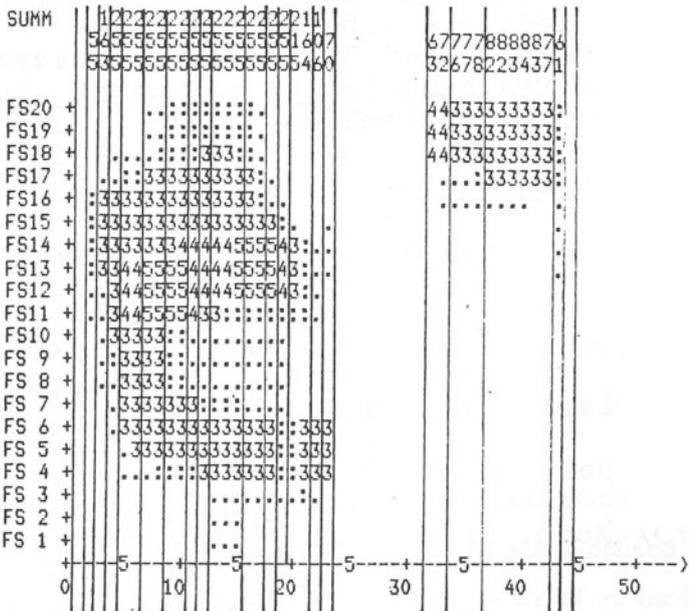
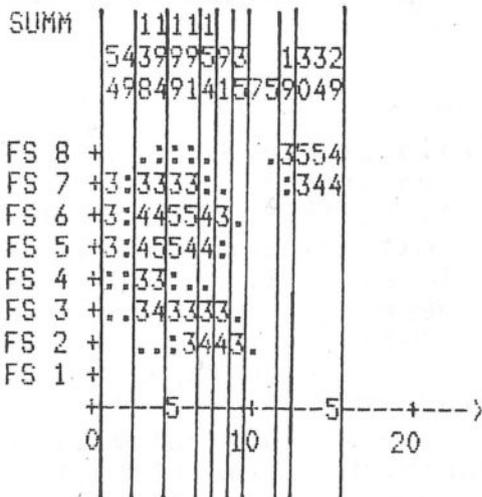


Bild 3: Darstellung der Primärbeschreibung, Vorverarbeitung und Ergebnisse des Clusters (senkrechte Linien Clustereinteilung)

### 3. Beschreibung der Experimente

Zunächst galt es einen Clusteralgorithmus zu finden, der zu der genannten unscharfen Beschreibung entsprechende Cluster liefert. Dabei wurde auf den von Mann /5/ entwickelten Algorithmus zurückgegriffen. Erste Untersuchungen zeigten, daß dieser Algorithmus nur dann geeignet ist, wenn es sich bei den Eingangsgrößen um quasistationäre Zustandsbeschreibungen handelt. Die in der Sprache typischen Gradientenverläufe werden von dem Algorithmus nicht als solche erfaßt. Der Algorithmus arbeitet zunächst auch formal ohne Hintergrundwissen. Bei Veränderung des Experimentiersystems lassen sich aber bestimmt noch Verbesserungen erzielen. So galt das Hauptaugenmerk der bisherigen Experimente der Vorverarbeitung. Sie muß so arbeiten, daß sie aus der zeitdiskreten, spektralen Beschreibung des Sprachsignales solche Zustände so herausfiltert, daß eine Zustandsfolgebeschreibung für das eingehende Sprachsignal ermittelt wird. Diese als Eingangsgröße für den Clusteralgorithmus verwendet, liefert recht ermutigende Ergebnisse.

Als Testwortschatz wurde ein Ziffernwortschatz verwendet, der in der Arbeitsgruppe erarbeitet wurde und von dem bereits Vergleichsergebnisse hinsichtlich des Aufwandes und der Erkennungsraten mit anderen Erkennungsalgorithmen vorliegen. Dadurch kann eine Aussage über die Leistungsfähigkeit des Systems erbracht werden. Für die Vorverarbeitung wurden verschiedene Softwaremodule eingesetzt, die in der Arbeitsgruppe im Rahmen anderer Projekte bereits entstanden waren. Dabei handelt es sich in erster Linie um Filter- und Normierungsmodule. In den Experimenten wurden zwei verschiedene Verarbeitungswege untersucht (Bild\_3). Der erste Weg wird besprochen, um eine hohe Datenkompression zu erreichen. Dies geschieht linear in Zeit- und Frequenzrichtung. Der zweite Weg dient zu Untersuchungen inwieweit formale Algorithmen eine Hervorhebung relevanter Merkmale erzeugen können. Dies geschieht durch die Anwendung statistischer Filter und mehrerer Normierungen. Diese stoßen bereits sehr stark an die Grenzen der Leistungsfähigkeit unserer Rechentechnik, so daß die Bearbeitungszeiten sehr hoch liegen.

Eine Weiterverarbeitung der entstehenden Ausgangssignale ist mit der Komponente "URI" des Expertensystems EXA 1630 /1/ möglich. Dieses Programm ist in der Lage, Zugehörigkeitsfolgen zu Symbolfolgen zu verarbeiten.

### 4. Literatur

- /1/ Langmann, D.: EXA 1630 - Experttechnologie zur Sprachsignal-  
auswertung mit unscharfem Zugang  
TU Dresden, Dissertation A, 1990, eingereicht.
- /2/ Langmann, D.: Sprachsignalauswertung mit dem Expertensystem  
EXA 1630, Beitrag in diesem Band.
- /3/ Steinhagen, H.E., Fuchs, S.: Objekterkennung  
Verlag Technik Berlin, 1976.
- /4/ Boklisch, S.F.: Prozeßanalyse mit unscharfen Verfahren  
Verlag Technik Berlin, 1987.
- /5/ Mann, S.: Ein Lernverfahren zur Modellierung zeitvarianter  
Systeme mittels unscharfer Klassifikation  
TH Karl-Marx-Stadt, Dissertation A, 1984.