

Automatische Spracherkennung

Heute möchte ich Ihnen über ein sehr interessantes und perspektives Thema erzählen.

Spracherkennung oder auch automatische Spracherkennung ist ein Teilgebiet der angewandten Informatik, der Ingenieurwissenschaften und der Computerlinguistik. Sie beschäftigt sich mit der Untersuchung und Entwicklung von Verfahren, die Automaten, insbesondere Computern, die gesprochene Sprache der automatischen Datenerfassung zugänglich macht.

Die Forschung an Spracherkennungssystemen begann in den 60er Jahren, verlief damals allerdings erfolglos: Die von privaten Firmen entwickelten Systeme ermöglichten unter Laborbedingungen die Erkennung von einigen Dutzend Einzelwörtern.

Erst Mitte der 80er Jahre kam die Entwicklung weiter voran. In dieser Zeit entdeckte man, dass man durch Kontextprüfungen Homophone unterscheiden konnte. Indem man Statistiken über die Häufigkeit bestimmter Wortkombinationen erstellte und auswertete, konnte man bei ähnlich oder gleich klingenden Wörtern entscheiden, welches gemeint war. Diese so genannten Trigrammstatistiken wurden anschließend ein wichtiger Bestandteil aller Spracherkennungssysteme. 1984 stellte IBM ein erstes Spracherkennungssystem vor, das etwa 5.000 englische Einzelwörter erkennen konnte. Das System brauchte für einen Erkennungsvorgang jedoch mehrere Minuten Rechenzeit auf einem Großrechner.

Und nur 1991 stellte IBM erstmals auf der CeBIT ein Spracherkennungssystem vor, das 20.000 bis 30.000 deutsche Wörter erkennen konnte.

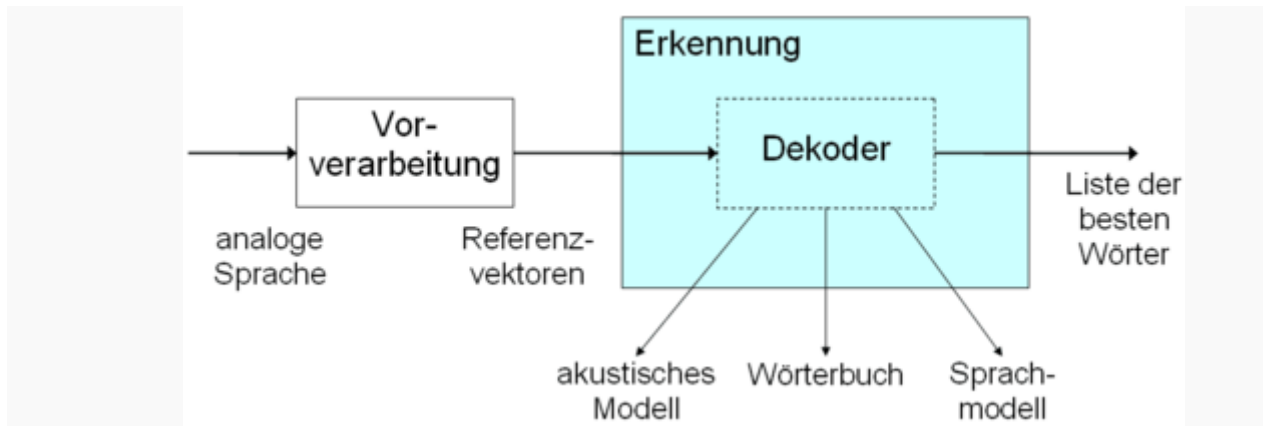
Heute kann grob zwischen zwei Arten der Spracherkennung unterschieden werden:

1. sprecherunabhängige Spracherkennung und
2. sprecherabhängige Spracherkennung.

Charakteristisch für die sprecherunabhängige Spracherkennung ist die Eigenschaft, dass der Benutzer ohne eine vorhergehende Trainingsphase sofort mit der Spracherkennung beginnen kann. Der Wortschatz ist jedoch auf einige tausend Wörter begrenzt.

Sprecherabhängige Spracherkenner müssen vom Benutzer vor der Verwendung auf die eigenen Besonderheiten der Aussprache trainiert werden. Ein Einsatz in Anwendungen mit häufig wechselnden Benutzern ist damit nicht möglich. Der Wortschatz ist im Vergleich sehr viel größer als der der sprecherunabhängigen Erkenner.

Bei beiden Arten wird jedoch derselbe Algorithmus gebraucht.



Ein Spracherkennungssystem besteht aus folgenden Bestandteilen: Einer Vorverarbeitung, die die analogen Sprachsignale in die einzelnen Frequenzen zerlegt.

Die Vorverarbeitung besteht im Wesentlichen aus den Schritten:

Abtastung,

Filterung,

Transformation

des Signals in den Frequenzbereich und Erstellen des Merkmalsvektors.

Bei der Abtastung wird das diskrete Signal digitalisiert, also in eine elektronisch verarbeitbare Bitfolge zerlegt, um es einfacher weiterverarbeiten zu können.

Die wichtigste Aufgabe des Arbeitsschrittes Filterung ist die Unterscheidung von Umgebungsgeräuschen wie Rauschen oder z. B. Motorengeräuschen und Sprache. Dazu wird zum Beispiel die Energie des Signals oder die Nulldurchgangsrate herangezogen.

Für die Spracherkennung ist nicht das Zeitsignal, sondern das Signal im Frequenzbereich relevant. Dazu wird es mittels FFT transformiert. Aus dem Resultat, dem Frequenzspektrum, lassen sich die im Signal vorhandenen Frequenzanteile ablesen.

Anschließend findet die tatsächliche Erkennung mit Hilfe akustischer Modelle, Wörterbücher und Sprachmodellen statt.

Heute gibt es eine Menge von Spracherkennungsprogramme wie bzw.: Stiftung Warentest, OS X Lion, Siri, Dragon Naturally Speaking, VoiceNavigator und andere mehr.

Diese Software wird in der Medizin, in der Sicherheitstechnik, in Telefongesellschaften, in den nationalen Sicherheitsdiensten verwendet. Gerade in der Medizin werden zunehmend Spracherkennungssysteme bei der Erstellung von

Befunden und Arztbriefen eingesetzt. In Telefon nutzen interaktive Anwendungen zunehmend die automatische Erkennung und Sprachsynthese.

Dieses Gebiet hat noch viele Richtungen und Perspektiven für die Entwicklung.