

# Computer lernen hören

Weil Rechner immer besser SPRACHE ERKENNEN, könnten Tastatur und Maus als Bedienungselemente schon bald ausgemustert werden

VON NIELS BOEING

Computer verlangen Handarbeit: Tagein, tagaus hacken Millionen Menschen wie Kulis mit gebeugtem Rücken auf die Tastatur, um Rechner mit Daten zu füttern. Schieben die Maus über ein Plastiktablett, um Symbole anzuklicken. Die meisten hatten sich das Informationszeitalter eher so vorgestellt: entspannt mit einem Drink in der Hand im Sessel sitzen und dem Rechner eher nebenbei zuzuhören, welche Website er suchen soll. Zwischendurch E-Mails und Excel-Tabellen schreiben, ohne einen Finger zu krümmen.

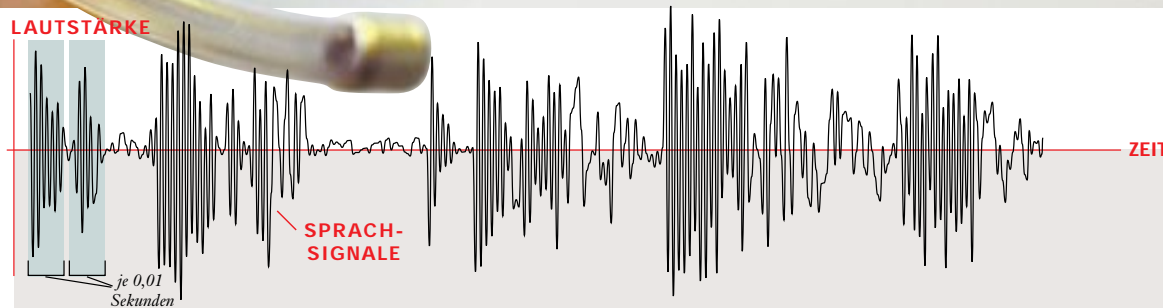
Nun, das mit dem Drink ist vielleicht ein wenig übertrieben. Der Rest jedoch wird bald Routine sein. Denn die Anstrengungen der Software-Industrie, den Computern Ohren zu machen, sind weit gediehen – Experten erwarten in diesem Jahr den Durchbruch der Spracherkennung als Standardtechnologie zur Computersteuerung. Sicherstes Indiz: Chip-Gigant Intel will künftig Schaltungen, die die Rechenoperationen zur Spracherkennung abwickeln, direkt auf den Prozessoren integrieren.

Noch vor wenigen Jahren beschränkte sich die Spracherkennung in der Alltagspraxis auf die Zahlen eins bis zehn. Das half zum Beispiel Verbrauchern, sich durch das Labyrinth einer Telefon-Hotline zu hangeln. Inzwischen hören sogar PCs aufs Wort. „Überall, wo ich reintippen kann, kann ich auch reindiktieren“, sagt Klaus Stanglmayr, Produkt-Manager von Philips Speech Processing.

Philips' Freespeech 98 gehört mit IBMs ViaVoice98 und Dragons NaturallySpeaking zur neuesten Generation komplexer Spracherkennungs-Software, die auch PCs das Hören beibringt (Mindestausstattung: 200-Megahertz-Prozessor und 48 Megabyte Arbeitsspeicher). Mit ihrer Hilfe lässt sich das Betriebssystem Windows per Zuruf steuern. Auf das Kommando „in Word diktieren“ öffnet sich prompt die Textverarbeitung, und das folgende Diktat – in ganz normaler Sprechgeschwindigkeit – wird automatisch als Word-Dokument angelegt.

Weniger komplexe Spracherkennungs-Software dürfte den ohnehin boomenden Absatz von Kleincomputern bald kräftig in die Höhe treiben. Dort geht die Entwicklung längst zur Verschmelzung von Palmtop, digitalem Adressbuch und Handy. Nur lässt sich die Tastatur, die man zum Schreiben für E-Mails und Telefonnummern braucht, nicht beliebig verkleinern. Die Geräte sollen aber in jede Jackentasche passen. Was liegt also näher, als hier zur Eingabe per Sprache überzugehen? Schon gibt es die ersten Handys, die zehn gesprochene Namen speichern können und bei Aufruf die entsprechenden Telefonnummern selbst wählen.

Was dem privaten Nutzer die Bedienung erleichtert, erspart der Dienstleistungsbranche handfeste Kosten. Wozu den Operator eines Call Centers



Tausende von menschlichen Stimmfrequenzen überlagern sich zu einem SPRACHSIGNAL. Wird es digitalisiert, kann ein Spracherkennungsprogramm in drei Schritten ermitteln, was gesprochen wurde:

- 1 Das Signal wird in kleine Abschnitte von jeweils einer Hundertstelsekunde Länge zerlegt. Mit einem speziellen Verfahren berechnet die Software Lautstärken und Schwankungsraten der verschiedenen Frequenzen. Die werden kategorisiert in 16 Tonhöhen, die charakteristisch sind für das menschliche Hörvermögen, und als Diagramm gespeichert.
- 2 Nacheinander werden diese dann mit den Diagrammen

## Vom Signal zum Wort

aller elementaren Laute (Phoneme) der deutschen Sprache verglichen, die im Lautarchiv der Software festgehalten sind. Ergebnis ist eine Folge von Phonemen, beispielsweise „d-u-sch-p-i-n-s-t-o-t-a-l“.

- 3 Welche Wörter zu dieser Lautfolge passen, wird mit dem so genannten Trigramm-Sprachmodell bestimmt. In dem sind die Wahrscheinlichkeiten aller erdenklichen Kombinationen aus drei Wörtern des Software-Vokabulars festgehalten. Gewonnen wurden sie aus unzähligen deutschen Sprachproben. Daraus ergibt sich, dass die Wortfolge „du spinnst total“ viel wahrscheinlicher ist als „dusch bin es tot Aal“.

ein Stichwort eintippen lassen, wenn ein Computer am anderen Ende der Strippe selbst in der Datenbank nachsehen kann? Vorreiter ist hier seit zwei Jahren die vollautomatisierte Zugauskunft der Schweizerischen Bundesbahnen (SBB). Seit September operiert ein weiteres automatisches Call Center: ALF, die „Aktuelle Lufthansa Fluginformation“. Das erste automatische Telefonbuch – für die Stadt Aachen – befindet sich derzeit in der Testphase.

Dass die Spracherkennung keine Utopie mehr ist, haben vor allem immer leistungsfähigere Chips und billigere Datenspeicher ermöglicht. Denn die digitalen Ohren sind wahre Byte-Fresser. Je mehr verschiedene Wörter sie verstehen sollen, desto mehr Speicherplatz und Rechenleistung benötigen sie.

Ich möchte von Berlin nach Hannover fahren – dieser Satz ist für den Computer zunächst nichts als Lärm,

den er, wie gewohnt, digitalisiert und somit umwandelt in eine Folge von Nullen und Einsen. Aus welchen Lauten der Klangbrei besteht, ermittelt die Software dann durch Vergleich mit gespeicherten Referenzlauten (siehe Kasten). Die entstammen einem „Durchschnittsdeutsch“, das anhand von 220 Sprechern verschiedener Mundarten und Akzente aus Deutschland, Österreich und der Schweiz entwickelt wurde.

Hat der Rechner die Lautfolge erkannt, ist er allerdings noch nicht schlauer als ein Tourist, dem ein Marktschreier in Nairobi einen Satz in Suaheli entgegenschleudert. Denn unklar ist, wo die einzelnen Wörter anfangen oder aufhören. Die Software stellt deshalb „Worthypothesen“ auf, die von einem so genannten Trigramm-Sprachmodell ausgewer-

tet werden. Die wahrscheinlichste Version wird auf dem Bildschirm angezeigt oder als Grundlage einer Datenbanksuche verwendet.

Inzwischen existieren Spracherkennungsprogramme für zehn Sprachen. Am leichtesten ließ sich Englisch entwickeln. „Im Chinesischen hingegen musste die Tonhöhe der Silben mitmodelliert werden“, sagt Christian Poyer, Leiter der Sprachtechnologie bei IBM. Deutsch hat ein anderes Problem: Da die Software bislang keine Grammatik berücksichtigen kann, müssen „Baum“ und „Bäume“ als getrennte Wörter abgelegt werden, der Speicher für den Wortschatz bläht sich auf.

Das Erkennen von elementaren Lauten (Phonemen) und von Worten läuft in den heutigen Programmen gleichzeitig in einem einzigen Rechenprozess ab. Je kleiner hierbei der Wortschatz ist, desto mehr Ausspra-

chevarianten können bei den Phonemen berücksichtigt werden. Bei einem kleinen Vokabular von bis zu 2000 Wörtern ist die Erkennung daher schon unabhängig vom Sprecher und in automatischen Call Centern einsetzbar. Damit auch die PC-Programme mit ihrem aktiven Wortschatz von über 60 000 Wörtern unabhängig vom einzelnen Sprecher funktionieren, wäre dagegen eine Rechnerleistung nötig, die heute noch kein PC bringt.

Deshalb benötigen sie bislang eine Trainingsphase. Geduldig wie ein Märchenonkel muss der Nutzer seinem Computer Mustertexte vorlesen. Die Software passt ihre schematisch eingestellte Aussprache an seine sprachlichen Eigenheiten an. Je mehr Trainingsstunden investiert werden, desto höher ist die Erkennungsrate, bis der Rechner im Idealfall wie ein Hund aufs Herrchen hört – aber auf niemanden sonst mehr. Eines der nächsten Ziele der Entwickler liegt daher darin, auch Programme mit großen Wortschatzen unabhängig vom Sprecher zu machen.

Ein anderes verfolgen sie, um den Sinn gesprochener Wörter zu verstehen. „Keines der jetzigen kommerziellen Systeme kann aber Sprache wirklich analysieren und einen Kontext verstehen“, sagt Wolfgang Wahlster. Der Künstliche-Intelligenz-Forscher aus Saarbrücken leitet das Forschungsprogramm Verbomobil, das genau dieses Problem knacken will, um gesprochenes Deutsch per Computer sinnvoll und verlässlich ins Englische oder Japanische zu übersetzen.

Die im Dezember vorgestellte Version Verbomobil II kann Sätze grammatikalisch analysieren, um verschlungene Endungen zu ergänzen. Sie erkennt den Kontext über mehrere Sätze hinweg: So wird das Wort „Vorstellung“ mit *performance* übersetzt, wenn von einem Theaterstück die Rede ist, aber mit *idea*, wenn jemand sagt, er habe keine Vorstellung, wie alt der Schauspieler sei. Highlight der Software sei die „automatische Protokollierung“, so Wahlster. „Die Software kann ein Gespräch auf das Wesentliche verdichten und so ein schriftliches Ergebnisprotokoll erzeugen.“

Doch die neue Spracherkennungstechnologie wird nicht jeden zufrieden stellen. Das SBB-Auskunftssystem benötigt nur einen halben Arbeitsplatz für den Wartungstechniker – als herkömmliches Call Center hätte es 137 Jobs geschaffen.

Wer mit literarischer Sprachgewalt in den heimischen PC diktieren will, wird sich über die „Kreativität“ der Software nur wundern. Die macht aus „die phlegmatisch Eingeeölnen, die Verkorksten und Vernölten“ von Max Goldt „die Pflicht thematisch einen wütenden tiefer Corps dann unvermittelt in“. Ein IBM-Bericht zur hauseigenen Spracherkennungstechnologie formulierte es so: „Tests mit Texten des ‚Titanic‘-Kolumnisten Max Goldt brachten durchweg unglückliche Ergebnisse.“

### ANGESPITZT

„Erektionsfähigkeit ohne gleichzeitiges Begehren ist ein ganz und gar sinnloser Zustand“

GÜNTER AMENDT

Soziologe und Sexualforscher, über den Drang, eine Psycho-Droge zu entwickeln, die die körperlichen Wirkungen von Viagra unterstützt

„Manche sozialdemokratischen Amtsträger reden ganz offen darüber, wie man die Studiengebühren nennen muss, damit man sie einführen kann“

KLAUS LANDFRIED

Präsident der Hochschulrektorenkonferenz

„Es wäre fatal, unter dem Deckmantel der Ökologie Fremdenfeindlichkeit zu schüren“

JOSEF REICHHOLF

Zoologe, zur Trennung von „heimisch“ und „fremd“ in der ökologischen Diskussion über eingeschleppte Tier- und Pflanzenarten

„Das Alte Testament war das erste Lehrbuch der Genetik“

STEVE JONES

Genetiker

### WOCHENSPIEGEL

#### BRUTPFLEGE

ALTERSFORSCHUNG: Frauen mit vielen Kindern haben eine geringere Lebenserwartung als solche mit wenigen oder ohne Kinder. Eine englische Studie bestätigt damit eine Annahme der Evolutionsbiologie: Die anstrengende Aufzucht von Nachkommen verhindert die Gesunderhaltung des eigenen Körpers.

#### WIRBELSTÜRME

WETTERVORHERSAGE: Nord- und Südamerika drohen in diesem Jahr unnatürlich viele Hurrikane. Auf Grund von veränderten Meeresströmungen prognostiziert die Colorado State University 14 schwere Stürme – doppelt so viele wie im langjährigen Mittel.

#### VERSCHREIBUNGEN

MEDIKAMENTE: Deutschlands Ärzte haben 1997 im Vergleich zum Vorjahr 11,3 Prozent weniger Rezepte ausgestellt, zeigt der jüngste Arzneiverordnungs-Report. Weil die Ärzte aber mehr teure Spezialpräparate verschrieben, sank der Arzneimittelumsatz nur um 1,7 Prozent.

#### STRAHLUNG

ATOMMÜLL: Radioaktiver Abfall aus unterirdischen Endlagern verbreitet sich schneller als erwartet. US-Forscher entdeckten, dass natürliche wasserlösliche Partikel die wasserunlöslichen radioaktiven im Grundwasser verteilen.

