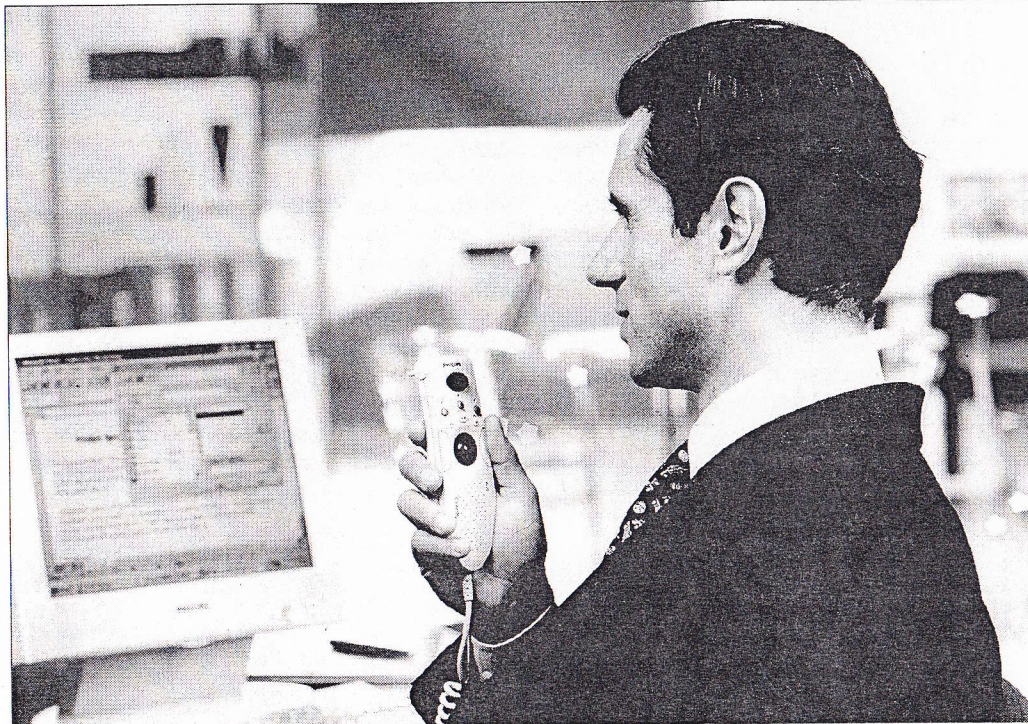


Wenn der Computer zum Diktat bestellt ist

Große Fortschritte bei der Spracherkennung – Einsatz auch bei Mobiltelefonen und Videorecordern

Der HAL 9000 ist ein Supercomputer, der alle Funktionen eines Raumschiffes selbstständig steuern kann und noch eine weitere bemerkenswerte Fähigkeit besitzt: Er versteht Sprache. Was wenige wissen: Wenn man statt H-A-L die im Alphabet jeweils folgenden Buchstaben zusammensetzt, ergibt sich der Name eines der weltweit größten Computerherstellers: IBM. Ein Zufall? Wohl kaum.

„Wir forschen seit Jahrzehnten im Bereich elektronischer Spracherkennung“, so Wolfgang Karbstein, Leiter des Geschäftsbereichs Spracherkennung bei IBM in Stuttgart. Mit „ViaVoice 98 Home“ hat IBM kürzlich ein Spracherkennungs-Kit auf den Markt gebracht, das es auch Heimanwendern ermöglicht, den Computer zum Diktat zu rufen. „Über ein Headset wird der Text entweder direkt in Micro-



Maus, Mikrofon und Lautsprecher enthält SpeechMike, das damit die Funktion eines Diktiergerätes erfüllt.

FOTO: GMS

soft Word 97 diktiert oder in das ViaVoice Speakpad gesprochen und danach in die bevorzugte Textverarbeitung übertragen“, erklärt Karbstein die Funktionsweise. Headsets sind Kopfhörer-Mikrofon-Kombinationen, wie sie zum Beispiel auch in Call-Centern oder von Piloten im Cockpit getragen werden.

Wie aber kann es sein, dass eine Maschine menschliche Sprache versteht und in einen Text umwandelt? „Das Programm versucht zunächst die Phoneme zu erkennen“, erklärt Karbstein – also die kleinsten Teile der Sprache, deren Austausch einem Wort eine andere Bedeutung geben würde. Aus den Phonemen werden dann Silben identifiziert und zu einem Wort zusammengesetzt. „Durch eine sogenannte Trigramm-Statistik ermittelt ViaVoice 98 dann die Wahrscheinlichkeit des identifizierten Wortes, wobei die vorher und anschließend gesprochenen Worten berücksichtigt werden. Das grenzt schon an höhere Mathematik.“

setzen aufschreiben zu müssen“, sagt Thomas Eisele, Leiter der Philips-Forschungsgruppe Spracherkennung in Aachen. Auf einer im Lieferumfang enthaltenen CD-Rom sind 13 Sprachen gespeichert, in denen diktiert werden kann, darunter Englisch und Französisch, aber auch Flämisch und sogar österreichisches Deutsch. „Im Sprachlexikon sind die Worte in Lautschrift abgelegt“, erklärt Eisele. „Wenn die Phonemfolge einem bestimmten gespeicherten Wort entspricht, wird dieses auf dem Computerbildschirm ausgegeben.“

Intelligente Telefondienste

Die Vorteile der Spracherkennung beschränken sich jedoch nicht auf reine Sekretariatstätigkeiten. „Diese Funktion wird unseren Alltag revolutionieren“, ist Thomas Eisele sich sicher. „Wir haben bereits ein Mobiltelefon auf den Markt gebracht, das per Sprache bedient werden kann. Diese Technologie ließe sich beispielsweise aber auch auf die Bedienung eines Videorekorders

64 000 Wörter umfasst der mitgelieferte ViaVoice 98-Wortschatz. Er ist aufrüstbar auf bis zu 128 000 Wörter. Um die Trefferquote von ViaVoice 98 zu erhöhen, kann das Programm auch trainiert werden: Bis zu 256 vorgegebene Sätze spricht der User dazu in seinen PC, das Programm analysiert die Sprechweise und passt die Spracherkennung der Artikulation des Sprechenden an.

Weg von Tastatur und Maus

Mit einem vergleichbaren System arbeitet Philips in seiner Spracherkennungssoftware „Freespeech 2000“. Statt eines Headsets kann der Käufer auch ein sogenanntes SpeechMike benutzen, welches die Funktion von Maus, Mikrofon und Lautsprecher in sich vereinigt. Die Nutzung des Geräts erinnert entfernt an die eines Diktiergeräts.

„Wir wollten weg von einer Bedienung über Tastatur und Maus hin zu einem Gerät, das in der Anwendung dem entspricht, was FreeSpeech 2000 tatsächlich ist: ein intelligentes Diktiergerät, mit dessen Hilfe Mitarbeiter davon entlastet werden, stupide die gesprochenen Sätze ihres Vorge-

übertragen. Viele weitere intelligente Telefondienste sind denkbar, etwa Wettervorhersagen, Fahrplan-Informationen oder Staumeldungen. Man ruft eine bestimmte Service-Nummer an und spricht einfach die gewünschte Auskunft in den Hörer. Der Computer liest dann die gefundenen Informationen vor.“

Ein Beispiel der Möglichkeiten für den Einsatz von Spracherkennung in der Zukunft bietet auch das European Media Laboratory (EML) in Heidelberg mit seinem Deep Map Informationssystem. „Deep Map ist ein mobiles Sprachübersetzungssystem, in das ein Touristenführer integriert ist“, so Projektleiter Rainer Malaka. „Wie sah das Schloss aus, bevor es zerstört wurde? Was ist das da vorn für ein Gebäude? Wo kann ich Karten für das Konzert bekommen? Auf all diese Fragen weiß Deep Map eine Antwort.“ Ziel sei es, den Gebrauch von Spracherkennung so selbstverständlich werden zu lassen, dass es dem Nutzer nicht mehr auffällt, ob er mit einem Menschen oder einer Maschine spricht.

Venio Piero Quinque, gms