03.09.2002 - Computer finden Worte

Sprechende Computer kennt man aus zahlreichen Filmen. Legendäres Beispiel ist der Bordcomputer "HAL" in Stanley Kubricks Film "2001 Space Odyssee". Sanft, nüchtern und perfekt ist seine Stimme. Im realen Jahr 2002 sieht das ganz anders aus. Sprachlich stecken die Computer noch in den Kinderschuhen.

Abgehackt und hölzern klingen die Stimmen der Maschinen. Kaum verständliche Sätze ertönten noch vor einiger Zeit aus den Lautsprechern. Doch die Branche im Bereich der so genannten ‚Sprachsynthese' entwickelt sich. Immer mehr verbesserte Sprachsysteme drängen auf den noch neuen Markt. Kein Wunder, denn das Anwendungsspektrum hat sich in den letzten Jahren beträchtlich erweitert. Neben den computerisierten Auskunftsdiensten, z.B. bei der Rufnummernansage und den Reisehotlines geht es immer häufiger um Transaktionsservices bei Hotelreservierungen, Flugbuchungen und anderen Diensten.

Eine Vorreiterrolle bei dieser Entwicklung nimmt das **Institut für Kommunikation und Phonetik** (IKP) der Universität Bonn ein. Seit Jahren versuchen Ingenieure und Wissenschaftler, Computern das Reden beizubringen. Kein leichtes Unterfangen, weiß IKP-Professor Wolfgang Hess. „Unsere Arbeit dreht sich dabei stets um Verständlichkeit und Natürlichkeit." Trotz der hoch entwickelten Technik ist es den Wissenschaftlern bisher noch nicht gelungen, die Computerstimme menschlich und natürlich klingen zu lassen.

Die Bonner Forscher verfolgen den bereits 1988 entwickelten Ansatz der ‚non uniform unit selection' des japanischen Wissenschaftlers Yoshinori Sagisaka. Der Ansatz basiert auf dem Prinzip, sämtliche Laute innerhalb eines Satzes zu katalogisieren und mit einem Etikett zu versehen, das zusätzliche Angaben über die Länge oder Position im Satzgefüge speichert. Diese Datenbank enthält dann beispielsweise hundert Variationen des Lautes ‚au'. Nach dem ‚Baukastenprinzip' sucht der Computer anhand des Etikettes das dem Satzgefüge entsprechende ‚au' aus. Im Satz „Wie siehst du denn aus?" beispielsweise das ‚au' mit dem Etikett „Satzende, Frage".

Anhand dieser Vorgehensweise gelang es den Wissenschaftlern, einen Sprachautomaten zu entwickeln. Obwohl die Stimme noch nicht menschlich klingt, ist er fähig, betonend und verständlich vorzulesen. Um dem Computer Emotionen geben zu können, müssen jedoch noch Unmengen von weiteren Daten gesammelt werden. Selbst bei einem Vorleseautomaten, der sich so emotionslos wie ein Nachrichtensprecher anhört, müssen riesige Mengen an Daten verarbeitet werden. Denn einige Laute tauchen bis zu 14.000 mal in der Datenbank auf, und jeder Satz besteht aus Dutzenden bis Hunderten von Lauten.

Die Grundlagenforschung der Bonner Universität ebnete den Weg für die Computer-Industrie. Auf Grund der Forschungsergebnisse haben Weltkonzerne den Markt für Sprachsynthese entdeckt. Riesige Summen werden für die Entwicklung von Endprodukten für die Verbraucher ausgegeben. Angesichts dieser Budgets haben sich die Bonner Wissenschaftler aus der Forschung zur Entwicklung von Produkten für den Massenmarkt zurückgezogen. Nach Ansicht von Professor Hess stehen den Firmen deutlich mehr Mittel für die Forschung auf dem Gebiet ‚Voice Commerce' zur Verfügung.

Organisiert sind die Anbieter von so genannten ‚Voice-Commerce-Produkten' in einem neuen Branchenverband, dem ‚**Voice Applications Standards Committee e. V.**' (Vascom e.V.). Firmen wie IBM, Clarity Philips, Samsung - um nur einige zu nennen - gehören zu den Mitgliedern. Ziel des Verbands ist es, die Akzeptanz der natürlichsprachlichen Bedienung von Computern in der Industrie und bei der Bevölkerung zu fördern. Darüber hinaus plant der Verband die Standardisierung von Sprachdialogen und Schnittstellen. Profitieren von diesen Maßnahmen wird insbesondere der Verbraucher: Egal welchen Anbieter er anruft, das Frage- und Antwortspiel mit dem Computer soll gleich bleiben. Bereits in einigen Jahren soll nach Annahmen des Verbands der Markt für Voice-Commerce größer sein als E- und M- Commerce zusammen. So wundert es nicht, dass sich in den vergangenen zwölf Monaten die Technik im Bereich der Sprachsysteme sprunghaft weiterentwickelt hat. Unternehmen wie die **VoiceObjects AG** aus Bergisch Gladbach haben bereits erste Produkte mit hoher sprachlicher Qualität auf den Markt gebracht.

Mit ‚Natural Voices' bietet das Gladbacher Start-up Unternehmen ein System an, das jeden Text sicher und betont vorlesen kann. Außerdem lassen sich mit dem Softwarepaket intelligente Sprachdienste erstellen. „Das ist ein evolutionärer Markt, der sich entwickelt", prophezeit Tiemo Winterkamp, Mitgründer der VoiceObjects AG. Auskünfte, Reisebuchungen und andere Dienste, die heute im Internet zur Verfügung stehen, würden ohne technischen ‚Hickhack' per Sprache zugänglich. Zukünftig soll es möglich sein, mit den intelligenten Sprachdiensten fast wie mit einem Menschen zu reden. Rückfragen, z. B. bei dem Buchen einer Reise, werden das umständliche, sich durch das ‚Ja/Nein, 1/2/3 Menü-Hangeln' ersetzen. T-Venture, die im Rahmen einer Direktinvestition an VoiceObjects beteiligt ist, überzeugt vor allem das breite Einsatzspektrum der entwickelten Plattform, wie Investment Direktorin Michaela Tittmann erläutert: „Mit der von VoiceObjects realisierten Voice-Application-Plattform wird es für eine Vielzahl von Unternehmen interessant und vor allem wirtschaftlich attraktiv, sprachgesteuerte Dienste einzusetzen."

Neben den Unternehmensriesen gibt es auch noch zahlreiche andere Anbieter von Voice-Commerce Produkten. Das Bochumer Unternehmen **G Data** hat mit der vierten Version von ‚Logox' bereits ein hohes Maß an Qualität bewiesen. Die kostengünstige Software liest Web-Seiten, E-Mails oder beliebige Texte vor, erstellt E-Mails oder lässt den PC sogar telefonieren. Die Besonderheit des Programms ist, dass die Stimmauswahl variabel ist. Auch Dialekt kann verwendet werden.

Es ist nur eine Frage der Zeit, wann die Stimmen der Programme menschlich klingen. „Mit dem hohen Reifegrad von Sprach-Dialogsystemen vollzieht sich aktuell der Eintritt in einen Massenmarkt", stellt Luis Praxmarer, Vorstand der **Meta Group**, fest. Schon jetzt sind die Sprachsysteme in der Lage, unabhängig vom Sprecher Worte im Zusammenhang zu erkennen. Neben den klassischen Einsatzbereichen der Sprachsysteme in Auskunfts- und Informations-Systemen sieht Praxmarer bereits einen Trend hin zur Einbindung der Voice Applications in Unternehmensprozesse. „Sprache wird zum wichtigsten Steuerungsinstrument für den Zugriff auf Daten werden", ist auch Christoph Pfeiffer, Vorsitzender des Branchenverbandes Vascom e. V. überzeugt.

URL: http://www.media.nrw.de/special/know-how/2002/09/03/03089/

Vervielfältigung nur mit Genehmigung - 09.2002 © FTK