MIT'S MAGAZINE OF INNOVATION

TECHNOLOGY REVIEW

NR. 1 DEUTSCHE AUSGABE MAI 2003

Künstliche Organismen:
Forscher spielen Gott

Raketenflugzeuge:
Spritztour ins All

Quantencomputer:
Rechnen mit Unschärfe

Immobots:
Schlaue Systeme

Niels Birbaumer:
Der Gedankenleser

SPRACHE
NEUE CHANCE FÜR TELECOMS



VOICE FÜR TELECOMS

DIE UMTS-GEBEUTELTE TELEFONBRANCHE SETZT BISLANG VERGEBLICH AUF DATENDIENSTE. SIE SOLLTE SICH DESHALB AUF IHRE UREIGENSTE KOMPETENZ ZURÜCKBESINNEN – EIN NETZ AUS SPRACHE

VON THOMAS VAŠEK, ILLUSTRATION: ERIC TSCHERNE

Unter einer Telefonnummer in Bergisch-Gladbach bei Köln meldet sich die Zukunft. Man kann ihr sagen, wen man anrufen möchte, dann stellt sie einen durch. „Futura" arbeitet, derzeit noch im Teststadium, für die Deutsche Telekom, und sie ist eigentlich ganz sympathisch. Doch das freundliche Fräulein vom Amt ist nicht von dieser Welt: Ihre Stimme kommt von einem Computer, und der steht obendrein in London. Der von Futura angebotene Telefonbuch-Dienst ist, ehrlich gesagt, nicht besonders aufregend. Aber wie aufregend waren schon die ersten Seiten des World Wide Web, jener Erfindung, die innerhalb weniger Jahre zum Massenphänomen geworden ist? Ein neues WWW, eines auf der Grundlage von natürlicher Sprache, könnte ein ähnlicher Erfolg werden.

Es geht um eine Art Konterrevolution: Telekom-Anbieter für Festnetz wie Mobilfunk betrachten Sprach-Telefondienste fast als Altlast – solide genug, um noch ordentlich Geld in die Kasse zu spülen, aber ungeeignet, den Weg in die Zukunft zu bereiten und Investoren zu begeistern. Doch nach Jahren der Fixierung auf Daten, Daten, Daten könnte sich erweisen, dass die richtige Strategie in einer Verbindung der beiden Welten liegt – in der Verknüpfung von natürlicher, gesprochener Sprache („Voice") mit digitaler Informationsverarbeitung. Das Sprach-

netz könnte viele Dienste anbieten, die heute per Festnetz-Telefon, Handy oder WWW-Browser angesteuert werden: Produkte bestellen, Konten führen, Informationen einholen, Daten verwalten, E-Mails und SMS-Nachrichten lesen.

INTERFACE-STIMME

Eine simple Rechnung zeigt das Potenzial von Sprachanwendungen. Im Jahr 2002 nutzten 600 Millionen Menschen das Web. Doch über eine Milliarde Menschen haben ein Telefon – Tendenz steigend.

Drei Entwicklungsfäden könnten schon bald im Sprach-Netz zusammenlaufen. Erstens hat sich die Sprachtechnologie – dazu gehören Spracherkennung, die Synthese künstlicher Sprache ebenso wie Übersetzungssysteme, in den letzten Jahren weiterentwickelt. Nicht nur verbesserte Algorithmen, sondern vor allem auch gestiegene Prozessorleistung und Speicherkapazität erlauben heute effizientere Anwendungen (siehe Kasten „Spracherkennung").

Zweitens arbeitet die Branche bereits an einem Standard für das Netz aus Sprache. Er heißt VoiceXML und könnte für das Sprachnetz jene Bedeutung erlangen, die HTML für das World Wide Web hatte. Drittens wird der Trend zur Mobilität das Sprachnetz vorantreiben. Wer bloß zuhause oder im Büro vor dem Computer sitzt, ist mit Tastatur und Maus als Eingabemedien leidlich gut bedient. Doch mobile Dienste zwingen dazu, die Schnittstelle zwischen Mensch und Maschine neu zu überdenken: Im unverkabelten Leben könnte natürliche Sprache wieder zu Wort kommen. „Immer mehr Anwendungen verlangen eine natürliche Form der Interaktion – und die menschliche Stimme als wichtiges Interface," sagt David Nahamoo, Leiter der Sprachtechnologie-Abteilung der IBM Research Division.

Lieben Sie es, mit dem Stylus umständlich an Ihrem PDA herumzufummeln, während Sie an der Straßenkreuzung stehen? Würden Sie nicht lieber einfach Ihr Gerät fragen, wo Sie die nächste Pizzeria finden? Sprachtechnologie ist überall da nützlich, wo die Hände für wichtigere Dinge als zum Tippen gebraucht werden – zum Beispiel beim Autofahren, bei Operationen oder beim Inventar-Checken in einem Lager mit zehn Meter hohen Regalen.

Das Marktforschungsunternehmen Kelsey Group sieht die weltweiten Sprachtechnologie-Umsätze von 505 Millionen Dollar im Jahr 2001 auf zwei Milliarden Dollar im Jahr 2006 anschwellen. Das Wachstum bei den Kern-Sprachtechnologien soll der Studie zufolge einen Multiplikatoreffekt auslösen, der zu einer Umsatzsteigerung bei sprachgesteuerten Telefondiensten auf 27 Milliarden Dollar in den nächsten drei Jahren führen wird. „Voice-Applications sind die wichtigste Innovation seit der Erfindung der Maus und der grafischen Benutzeroberfläche für den Computer," meint gar Luis Praxmarer, Deutschlandchef des Marktforschungsunternehmens „Meta Group AG" und spricht ebenfalls von einem Umsatzvolumen für Voice Applications in zweistelliger Milliardenhöhe bis zum Jahr 2005.

Doch langsam: Hatten wir das nicht schon? Seit gut einem Jahrzehnt gilt Sprachtechnologie als Versprechen für die Zukunft. Inspiriert von Stanley Kubricks sagenhaftem Film-Computer HAL träumten viele von der Künstlichen Intelligenz, die uns versteht und – selbstverständlich – auch spricht wie ein Mensch. An simple Dinge wie sprachgesteuertes Telefonbanking und an Voice-Portale im Web haben wir uns mittlerweile gewöhnt – doch von einem Durchbruch konnte nicht gerade die Rede sein. Und mancher wird sich schon mal amüsiert oder geärgert haben über die Unbeholfenheit einiger Systeme.

Ausgereift ist die Technologie noch lange nicht. Kein System versteht jeden Akzent, und alle produzieren zumindest gelegentlich Unsinn. Immer noch lassen sich Spracherkennungs-Systeme leichter aus der Bahn werfen als die menschliche Telefonistin im Callcenter. Futura zum Beispiel, die nette Fräuleinstimme in Bergisch-Gladbach, hat einen Kollegen in Hannover, zu dem sie verbindet,

Spracherkennung: Evolutionäre Fortschritte

Automatische Spracherkennung ist ein mühsames Geschäft. Menschen sprechen undeutlich, husten, räuspern sich oder schieben mal ein „äh" dazwischen. Solche Eigenheiten machen Sprecherkennungsprogrammen ebenso zu schaffen wie jegliche Geräuschkulisse im Hintergrund. Der Fortschritt auf dem Gebiet war deshalb bislang eher evolutionär als revolutionär. Erst die letzten Jahre haben brauchbare Produkte wie etwa Diktiersysteme hervorgebracht. „Wir haben inzwischen eine neue Generation von Sprachérkennungssystemen, die mit spontaner Sprache, Versprechern, Unterbrechungen und Nebengeräuschen besser umgehen können," meint Wolfgang Wahlster, Sprachtechnologie-Experte an der Universität des Saarlandes.

Spracherkennung beginnt mit der Umwandlung von gesprochener Sprache in digitale Signale. Daraus bestimmt das Programm Phoneme, die elementaren Bausteine der Sprache, und kombiniert diese zu Wörtern. Da viele Wörter ähnlich klingen, braucht es jedoch Kontext. Spracherkennungssysteme berechnen dann die Wahrscheinlichkeit dafür, dass eine bestimmte Wortkombination vorliegt. „Der Fortschritt bei den Algorithmen ist konstant," sagt David Nahamoo von IBM. So habe man die Fehlerrate bei der Spracherkennung am Telefon in den letzten fünf Jahren immerhin um 25 Prozent reduzieren können. Die Fortschritte reichen für viele einfache Anwendungen.

Freilich: Die große Herausforderung der Forscher liegt weniger im Erkennen als im Verstehen von natürlicher Sprache („Natural Language Understanding", NLU). „Das Verstehen ist der schwierigste Teil," sagt Nahamoo: „Dazu muss man Bedeutungen und semantische Informationen modellieren. Ohne Semantik ist Verstehen nicht möglich." Verstehen braucht es beispielsweise bei der Übersetzung von gesprochener Sprache, „sonst kommt unweigerlich Unsinn heraus". Auf der diesjährigen CeBIT präsentierten die IBM-Forscher erstmals ein Echtzeit-Übersetzungssystem auf Basis von NLU, das drei zentrale Sprachtechnologien – Spracherkennung, Übersetzung und synthetische Sprachausgabe – miteinander kombiniert.



Kostenfaktor Callcenter: Sprachanwendungen ermöglichen Einsparungen in Milliardenhöhe.

wenn jemand „Restaurant Guide" sagt. Man kann die gewünschte Küchenrichtung angeben, hört dann kurze Beschreibungen und kann jederzeit „weiter" oder „reservieren" sagen. Wenn man sich entschieden hat und die Frage nach der Zahl der gewünschten Plätze mit „500" beantwortet, stürzt Futuras Kollege ab. Tiemo Winterkamp, Leiter Produktstrategie & Research der Bergisch-Gladbacher Firma Voice Objects und damit quasi Futuras Vater, ist der Absturz nicht so peinlich: „Das kann man feintunen." Er glaubt, dass immer mehr Sprach-Dienste entstehen werden: „Einfache Systeme wie Kontostandabfrage oder Flugbestätigung kann man heutzutage schon leicht realisieren."

Futura wurde nur als Demonstration gebaut, das echte Geschäft von Voice Objects ist Software für einen schwierigen Teil im Sprachnetz: Die Steuerung der Interaktion zwischen Mensch und Maschinen am Telefon. Denn noch können Computer wenig anfangen mit munterem Drauflosgeplapper à la „Tag, einmal nach Köln und zurück, Bahncard 2. Klasse, schnell bitte, der Zug kommt gleich." Sie brauchen die Infos Stück für Stück. Ein Dialog mit dem Computer wird nur so lebendig, wie er vorher programmiert wurde – Futura zum Beispiel lässt sich gerne unterbrechen, wenn sie ihren Namen hört, sagt dann eifrig „Jaa? Ich bin hier", und fängt wieder beim Hauptmenü an. An die Flexibilität von echten Menschen in der Schalterhalle oder in Call Centers reichen solche Voice-Dienste noch nicht heran. Andererseits sind sie ungleich schneller – Futura spricht heute schon mit hunderten Anrufern gleichzeitig. Schlangen wie am Bahnhof und Warteschleifen wie bei den Hotlines sind bei ihr und ihren synthetischen Kollegen nicht zu befürchten.

TREND-MOTOR-ANWENDUNGEN

Die Sprachtechnologie mag nicht perfekt sein. Doch immerhin funktioniert sie gut genug für eine Reihe von Anwendungen – und genau diese Anwendungen könnten sich als Motor für den Trend erweisen. „Jetzt geht es darum, Spracherkennung von einem reinen Tech-Spielzeug zu einer businessorientierten Technologie weiterzuentwickeln," sagt IBM-Forscher Nahamoo.

In Krisenzeiten locken vor allem Einsparungspotenziale. Studien zufolge bestehen rund 70 Prozent der Kommunikation zwischen Kunden und Unternehmen aus der Abfrage einfacher Informationen. Diese Kundenkontakte wickeln heute noch immer hauptsächlich Menschen aus Fleisch und Blut ab – und die kosten Geld. „Bis 2005 wird ein Großteil dieser Kommunikation via Telefon automatisch durch Sprachcomputer und Anbindung an das Internet realisiert," prognostiziert Luis Praxmarer von der Meta Group.

Ein Beispiel sind die sündhaft teuren, personalintensiven Callcenter. „Das ist ein 80-Milliarden-Dollar-Markt," sagt IBM-Mann Nahamoo: „Darin liegt eine Riesenchance." Nach Aussagen des Marktforschungsunternehmens Giga Information Group (USA) könnten die Kosten für Call Center und Hotlines gar um 20 bis 30 Prozent reduziert werden.

Jedes Netz braucht eine Sprache, in der sich Anwendungen, unabhängig vom jeweiligen Hersteller, untereinander verständigen können. Was HTML (Hypertext Markup Language) für das Web war, könnte der offene Sprachstandard VoiceXML (Voice eXtensible Markup Language) für das Sprachnetz werden.

Unter Führung von Branchenriesen wie AT & T, IBM, Lucent und Motorola unterstützen mittlerweile hunderte Firmen diesen Standard, der innerhalb des für Standards zuständigen World Wide Web Konsortiums (W3C) entwickelt wurde. HTML ermöglicht

jedermann, mit simplen Befehlen Webseiten zu gestalten. VoiceXML erlaubt es, Sprachanwendungen zu beschreiben. Während HTML von einem grafischen, in einen Rechner installierten Webbrowser ausgeht, steht im Zentrum von VoiceXML ein so genannter Voice-Browser mit Audio-Eingabe und -Ausgabe. Der Browser läuft nicht auf dem Gerät des Nutzers, sondern beim jeweiligen Plattformbetreiber – ein Voice-Browser kann hunderte Anrufer gleichzeitig bedienen. Das Sprachnetz ist indes keine Parallelstruktur zum Web, sondern es benützt dessen in den vergangenen Jahren gewaltig gewachsene Infrastruktur.

„Multimodalität" lautet das Zauberwort, das den entscheidenden Zusatznutzen für den Kunden bringen soll. Ein- und Ausgabe sollen per Sprache, aber eben auch über Tastatur oder Text möglich sein. Das schafft Flexibilität und Nutzerkomfort: Wer über Handy oder PDA sein Konto abfragt, will die Antwort vielleicht lieber als Textnachricht als von der Computerstimme erhalten. Die IBM-Entwicklungszentren Mannheim und Prag demonstrierten auf der diesjährigen CeBIT, wie der sprachgesteuerte oder auch schon multimodale Zugriff auf Webinformationen funktionieren könnte. Beispiel Stauinformationen: Die Anwendung auf



Handy-Kommunikation: Mobile Dienste erfordern Stimme als Schnittstelle.

VoiceXML-Basis läuft auf einem Server, zu dem mittels Telefon eine Sprachverbindung aufgebaut wird. Der Nutzer gibt die Abfrage per Sprache, etwa über Handy ein. Mit einem PDA könnte die Eingabe multimodal, also per Sprache oder Stylus, erfolgen. Der Server generiert dann die Ausgabe, wieder entweder in synthetisierter Sprache oder als Text auf dem Display beziehungsweise Bildschirm.

Erste VoiceXML-Anwendungen gibt es bereits. Im September 2001 richtete der US-Telefonriese AT & T eine Telefonauskunft für gebührenfreie Nummern auf VoiceXML-Basis ein – und spart damit immerhin etwa 20 Millionen Dollar pro Jahr für einschlägiges Personal.

Nicht jeder ist von VoiceXML so begeistert wie IBM-Forscher Nahamoo, dessen Unternehmen Teil des VoiceXML-Firmenkonsortiums ist. „Der Standard ist eigentlich schon wieder überholt," meint etwa Wolfgang Wahlster, Sprachtechnologie-Experte an der Universität des Saarlandes. Zwar sei VoiceXML für kleine Anwendungen gut geeignet und leiste damit einen Beitrag zur Popularisierung – jeder kann sich damit so leicht eine Voice-Site als aufgemotzten Anrufbeantworter bauen wie bisher eine private Web-Site. Für anspruchsvolle Sprachdialoge aber sei VoiceXML „ein Witz". Bereits im Oktober 2001 formierte sich eine unter anderen von Microsoft, Cisco und Intel unterstützte Konkurrenzveranstaltung namens SALT (Speech Applications Language Tags). Deren Spezifikation SALT 1.0 wurde dem W3C im August 2002 zur Prüfung vorgelegt.. Zugleich wird auch an einer mächtigeren VoiceXML-Variante gearbeitet. Ob VoiceXML oder SALT – letztlich wird sich





Echtzeit-Übersetzungssystem (IBM): Sprachtechnologie macht Fortschritte.

wohl ein Standard herauskristallisieren. Aus einer Vielzahl von Insellösungen lässt sich dann ein Netz weben.

Wann also beginnt der Boom? „Der Durchbruch kommt im Verlauf der nächsten zwei Jahre", sagt Jürgen Haas von der Sympalog AG i. I. Das i. I. steht für „in Insolvenz" und für die Aussicht auf einen Neuanfang. Was die Voice-Welt braucht, sei ein erfolgreiches Projekt bei einem großen Kunden.

DATENFIXIERUNG

Versuche hat es gegeben. Manche endeten als Peinlichkeit wie die erste Computer-Zugauskunft der Bahn: Ein Radiosender spielte einen Mitschnitt von einem Dialog vor, in dem der Computer den Anrufer zum Verzweifeln bringt. Die Sendung kursiert immer noch als Audiodatei im Internet. „15 000 Bahnhöfe – das ist natürlich auch ein brutal großer Wortschatz", sagt Haas. In der neuesten Version wiederum sei die Auskunft fast zu überzeugend: „Die Leute fangen an, dem System etwas zu erzählen, weil sie es gut finden. Es versteht sie aber nicht."

Ernüchtert sind Beobachter auch von der 01 90/05 55 55, unter der die Deutsche Telekom ziemlich unbeachtet ein Sprachportal mit Nachrichten, Kinoauskünften, Wetter und Aktienkursen anbietet.

Die Idee ist natürlich gerade für Telefonfirmen spannend. Mit relativ billigen Inhalten können sie das generieren, wovon sie fast nie genug haben können: Gebühren-Minuten. DSL hin, UMTS her – hier ist eine neue Einnahmequelle, für die beim Nutzer ein primitives Endgerät ausreicht. Traffic, Traffic, Traffic – eine hohe Auslastung des Netzes wird für Carrier umso wichtiger, je teurer die Infrastruktur ist.

Jeder der sechs UMTS-Lizenznehmer in Deutschland hat mindestens 8,4 Milliarden Euro dafür hingelegt. Schon dadurch wird der neue Mobilfunk-Standard zum teuersten aller Zeiten. „Den Telecom-Unternehmen ginge es eigentlich gar nicht schlecht – wenn nur die verdammt hohen UMTS-Kosten nicht wären", sagt Marc Ziegler, Managing Partner und Bereichsleiter Media/Online beim Beratungshaus Detecon. Die UMTS-Geschädigten setzen, sofern sie noch nicht aufgegeben haben, auf Datendienste. Gebannt verfolgen Analysten die Entwicklung von Quartal zu Quartal: Steigt der Anteil am Umsatz, der mit Datendiensten gemacht wurde wie geplant? Ja? Gut. Nein? Nicht gut.

„Wenn i-Mode kein Erfolg wird, sieht es düster aus", sagte im vergangenen Jahr der Chef des Mobilfunkers Vodafone D2 Deutschland, Jürgen von Kuczkowski. In diesem Frühjahr meldete der Betreiber E-Plus Zahlen für den aus Japan importierten Multimedia-Datendienst fürs Handy: Innerhalb von einem Jahr ließen sich nur 123 000 Kunden dafür gewinnen, geplant waren mindestens doppelt so viele – ein Ergebnis, das gut zu anderen Erfahrungen der Branche passt.

Eine Lektion haben die Telecoms schon gelernt: Nicht der WAP-Datendienst spülte Geld in die Kassen der Mobilfunkbetreiber – sondern der technisch triviale Kurznachrichtendienst SMS. Vielleicht ist es für die Telekommunikations-Branche wirklich besser, sich auf ihre alte, ja altmodisch wirkende Kompetenz zu besinnen. Und vielleicht stimmt ja, was die Marktforscher von Forrester Research prophezeien: Erst Sprachdienste würden den Weg der Telecom-Branche ins Daten-Zeitalter bereiten. Das gute, alte Telefon als Rettung für die Datenwelt: Das wäre ironisch genug. ⓣ