SCHWERPUNKT: AUTOMATISCHE SPRACHERKENNUNG VOR DEM DURCHBRUCH (?)

Spracherkennung – gestern, heute und übermorgen

Ahmet E. Çakir // Ergonomic Institut, Berlin

HIER LESEN SIE:

- wo die Technologie der Spracherkennung heute steht
- warum sich die Umwandlung von Sprache so schwierig gestaltet
- auf welchen (Spezial-)Gebieten solche Anwendungen sinnvoll sind

Die Umwandlung von Sprache in Text ist ein alter Menschheitstraum, an dem schon lange geforscht wird – von der „Eisernen Sekretärin" bis zu „Siri". Doch die vorhandenen Systeme schaffen es bis heute noch nicht, in allen Bereichen wirklich überzeugende Ergebnisse zu liefern. Die gesprochene Sprache und die geschriebene scheinen zu verschieden zu sein. Das bedeutet aber nicht, dass nicht auf speziellen Gebieten die maschinelle Spracherkennung erfolgreich eingesetzt werden kann. In manchen Fällen wird sie sogar die bestmögliche Anwendung bleiben oder werden. Welche das sind und künftig sein können, zeigen dieser und die weiteren Beiträge im Rahmen dieses Schwerpunkts.

„Sie werden morgen ein Labor besuchen, in dem sehr engagierte junge Forscher arbeiten. Sie sind sehr aufgeregt, weil sie vor dem Durchbruch stehen. Bitte sagen Sie nichts, die stehen nämlich immer vor dem Durchbruch. Wir stecken die neuen Mitarbeiter immer in dieses Labor. Nach zwei oder drei Jahren machen sie etwas Vernünftiges."

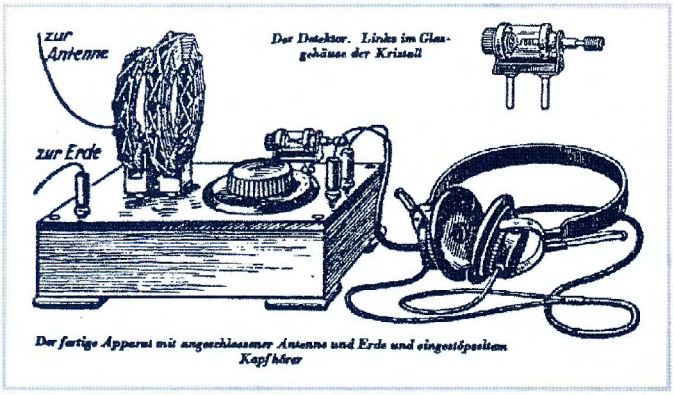
Diese Worte sprach der Leiter der Forschungslabors des damals wichtigsten EDV-Unternehmens der Welt, der mir im Jahre 1978 seine Arbeiten präsentierte. Und das Thema war Spracherkennung bzw. Sprachverarbeitung. Seitdem hat sich die Welt grundlegend gewandelt, aus EDV wurde IT. Vielleicht hat der Zahn der Zeit auch dieses Konzept, Informationstechnologie, arg angenagt. Heute schleppen Kinder Handys in der Hosentasche, deren Rechenpower größer sein soll als die komplette Kapazität der Mondlandemission Apollo 12, Rechner der Bodenstation in Houston eingerechnet. Wann ist der Durchbruch gelungen, der eigentlich bereits in den 1960er Jahren selbstverständlich gewesen sein sollte?

Begriffe und deren Hintergrund

Wenn man von Spracherkennung spricht, sollte man zunächst im Hinterkopf behalten, was man sich einst davon versprochen hatte. Ansonsten wird es schwierig sein, die Frage zu beantworten, wo wir mit der Technologie heute stehen. Das Wort Spracherkennung selbst weist auf keinen Anwendungszweck hin, außer dass „Sprache" und „Erkennen" darin verknüpft sind. Sie wird vermutlich auf ewig eine Baustelle bleiben wie die „Textverarbeitung", bei der man es mit Anwendungen wie „QuarkXpress" oder „Indesign" zwar viel weiter gebracht hat als mit der Schreibmaschine, aber das Ende der Fahnenstange noch im Nebel vermuten muss.

Eine Legende besagt, dass die erste Spracherkennungsmaschine, naturgemäß ohne diesen Namen, von einem der wich-

tigsten Erfinder der Geschichte der technischen Kommunikation, Guglielmo Marconi, erfunden worden wäre, und das in dem Jahr, in dem er den Physiknobelpreis erhielt. Die Maschine, die „Eiserne Sekretärin", würde einen Brief, den man hinein spricht, als Text auf Papier drucken. Allerdings soll Marconi zur Präsentation der Maschine vor der New Yorker Presse nicht erschienen sein.



Guglielmo Marconi war ein Pionier der drahtlosen Kommunikation ...

Wichtig bei dieser Legende ist die der Technik unterstellte Zielsetzung. Marconi, der die drahtlose Kommunikation erfunden und seinerzeit ein „Beinahe-Monopol" errichtet hatte, soll mit seiner Erfindung seine früheren Taten übertroffen und einem wichtigen Bedürfnis des damaligen Büros entsprochen haben: Sekretärinnen überflüssig machen. Diese waren zum einen sehr teuer, so man sie überhaupt finden konnte. Zum anderen bestand aber ein weiteres Problem in der Kommunikation mit dem „Sprecher" des Textes, dem Chef. Dieser will oder kann nicht immer dann arbeiten, wenn die Sekretärin kann oder will.

Die erste technische Lösung des Problems hatte Thomas A. Edison mit der Erfindung des Phonographen gefunden, den er eigens für die Aufgabe konstruiert hatte, die Kommunikation zwischen Chef und Sekretärin zeitlich zu entkoppeln.

Nunmehr sollte die „Eiserne" Sekretärin das Problem grundsätzlich lösen. Sinngemäß mit dem Ziel: „Vorne hineinsprechen – hinten Brief abholen!" Ähnliche Zielsetzungen schwirren auch heute in den Köpfen vieler umher, auch wenn man nicht nur Texte hinein spricht, sondern auch sonstige Daten eingibt. Ganz so für den Bedarf von Sekretärinnen von heute, für die die Texteingabe nur noch eine untergeordnete Rolle spielt. Wenn Maschinen in der Lage sind, ein Kopfschütteln oder Fingerschnippen richtig zu deuten, warum nicht etwas Gesprochenes?

Spracherkennung

Ernstzunehmende „Lösungen" für die Spracherkennung wurden ab den 1980er Jahren entwickelt. Man sprach von „akustischer Dateneingabe" und verstand darunter die Eingabe von Daten per Sprache in den Computer. Um der Lösung noch einen fertigen Anstrich zu verleihen, hat man auch ein Kürzel erfunden: CSE = Computer-Sprach-Eingabe. Man sprach von direkter Datenerfassung bzw. Kommunikation mit dem Rechner über CSE. Die CW (Computerwoche) vom Ende 1981 sprach von folgenden Vorteilen, die hierbei erreichbar wären bzw. bereits erreicht seien:

- optimale Anpassung an die vom Menschen als natürlich empfundenen Kommunikationsgewohnheiten,
- kurze Einarbeitungszeit,
- mehr Bewegungsfreiheit für den Benutzer beim Erfassen der Daten,
- vereinfachte direkte Datenerfassung,
- größere Sicherheit der Eingabe.

Fazit: „Die Spracheingabe bringt viele Vorteile gegenüber den traditionellen Eingabemöglichkeiten. Da das gesprochene Wort direkt zur Daten- oder Kommandoeingabe verwendet werden kann, werden Tätigkeiten des Ablesens und der wiederholten Übertragung von Daten mit ihren negativen Effekten der Ermüdung und der nachlassenden Konzentration beseitigt."

Hierzu sei angemerkt, dass der zitierte Bericht eine Erfolgsmeldung der CGK, Computergesellschaft Konstanz, auslegte, die sich im gleichen Jahr anschickte, auch die Schrifterkennung zu revolutionieren, heute OCR (Optical Character Recognition) genannt. Ihre Produkte waren über lange Jahre „state of the art", was unübertroffen bedeutet. Der Praxis bedeuteten aber beide Techniken nicht so viel, dass CGK ihre Muttergesellschaften (AEG und Nixdorf) lange überlebt hätte.

Eine Legende besagt, dass Marconi bereits Anfang des 20. Jahrhunderts die erste Spracherkennungsmaschine erfunden habe: die „Eiserne Sekretärin" ...

Im Laufe der nächsten zwei Jahrzehnte wurde aus der CSE etwas, was man nur noch mit englischen Kürzeln ausdrücken kann, aber nicht viel Neueres als Spracherkennung à la Marconi: Systeme, die das Gesprochene als Text erkennen und in digitalisierter Form drucken oder auf einem anderen Medium speichern.

Sprechererkennung

Andere, z. B. die Polizei, hatten davon diametral abweichende Ziele. Ihr ging es um die Erkennung des Sprechers und nicht des Gesprochenen. Bereits zu Beginn der 1980er Jahre schien die Sprechererkennung so weit gediehen zu sein, dass man von einem „akustischen Fingerabdruck" sprach. Wenn heute von Spracherkennung gesprochen wird, muss man fragen, ob es sich nicht um die Sprechererkennung handelt, also um eine biometrische Methode,

für die sich nicht nur die Polizei interessiert. Die Eigenschaften der persönlichen Sprache sind zum einen sehr viel alterungsunempfindlicher als z. B. das Aussehen, und zum anderen fälschungssicherer als ein Fingerabdruck (Anmerkung: Fingerabdrücke kann man zwar nicht fälschen, aber kopieren und wieder verwenden). Also das perfekte Mittel, um Personen, z. B. den Besitzer eines Handy oder eines Autos zu erkennen, der dann keinen Schlüssel mehr braucht. Oder seinen Zustand, in dem er nicht mehr Auto fahren darf.

Kommandoverarbeitung

Der Dritte im Bunde, die Kommandoverarbeitung, ist auch fast immer Teil des ersten, der Spracherkennung. Kommandosysteme sind Spracherkennungssysteme, die mit einem eingeschränkten Wortschatz oder gar mit einem künstlichen arbeiten. Ihr besonderes Merkmal ist die Abweichung von der natürlichen Sprache.

Sprachdialogsysteme, wie man sie heute – nicht immer zur großen Freude – bei Call-Centern oder großen Firmen erleben kann, wenn man dort anruft, sind meistens eine Kombination von Kommandosystemen mit Sprachausgabesystemen, die entweder eine Kombination von aufgezeichneten Wörtern oder eine synthetische Sprache ausgeben.

Die hier genannten Anwendungen haben sich sehr unterschiedlich bewährt. Unabhängig davon sind sie allesamt noch „entwicklungsfähig", was man auch negativ lesen kann, bzw. sollte, wenn man an eine Anwendung im eigenen Betrieb denkt. Der Grund des unterschiedlichen Erfolgs kann einerseits in der Aufgabe liegen, die das System bewältigen soll. Er kann aber auch an den Qualitätsmaßstäben liegen, die man an das Ergebnis stellt. Unabhängig von beiden muss man den Aufwand berücksichtigen, den man betreiben möchte, um ein solches System „zum Laufen" zu bringen.

Anwendungen und ihre Erfolgschancen

Spracheingabe und -erkennung

Die Erkennung natürlich gesprochener Sprache und deren Umwandlung in digitalisierte Form wird vollkommen unabhängig von der Technik eine unbefriedigende Baustelle bleiben. Dies hat zwar mehrere Gründe. Ausschlaggebend für diese Prognose ist aber die Tatsache, dass die gesprochene Sprache und die geschriebene nur selten gleich sind. Es ist ein viel verbreiteter Irrtum, dass ein gesprochener Text nur digitalisiert werden muss, um daraus geschriebenen zu erzeugen.

Die natürlich gesprochene Sprache hat eine andere Funktion für die Kommunikation als eine geschriebene. Während es sich bei einem geschriebenen Text um einen „Monolog" handelt, mit dem der Autor seine Inhalte ohne einen Eingriff seitens des Adressaten unterbreitet, wird derselbe Autor beim Sprechen die Natur der sprachlichen Kommunikation beachten. Beispielsweise ist die gesprochene Sprache mit Absicht zuweilen unpräziser, damit man ins Gespräch kommt.

Winner und Loser

Hinzu kommt, dass äußerst wenige Menschen in der Lage sind, „wie gedruckt" zu sprechen. Diese sind häufig in Berufen zu finden, deren Angehörige ohne ihre sprachlichen Fähigkeiten kaum erfolgreich sein können, so z. B. Ärzte, Anwälte oder Richter. Angehörige solcher Berufe waren bereits vor Jahrzehnten gute Kunden für eine Technik, die über die Spracheingabe funktionierte, Diktiergeräte. Für sie weisen heute verfügbare Techniken zusätzliche und erhebliche Vorteile auf, z. B. ein großes Wörterbuch mit bis zu einigen Millionen „Wörtern". Anders als die früheren Geräte, die auf Band speicherten und nur sequentiell als Sprache abzuhören waren, kann man seit Langem den gesprochenen Text fast zeitgleich am Bildschirm sehen. Navigieren im Text ist somit unvergleichlich schneller und einfacher geworden.

Die Angehörigen der Berufe, bei denen man viel mit natürlicher Sprache umgehen muss, werden aus diesem Grund auch von Spracherkennungssystemen profitieren, falls dies noch nicht der Fall sein sollte. Ein weiterer Grund kann in der Häufigkeit der Nutzung liegen, weil man effiziente Mechanismen entwickelt, um mit den fehlenden Fähigkeiten der Technik umzugehen. Wer dies nicht tut, gehört mit großer Wahrscheinlichkeit zu den Losern.

Was auch unsere Kinder noch für Zukunft halten werden

Hingegen wird die „Eiserne Sekretärin", gar eine solche, die einen in Deutsch gesprochenen Text in Englisch ausgibt, unabhängig von den Fähigkeiten der verfügbaren Technik allenfalls Vorlagen für unfreiwillige Komik liefern, wie man sich auf einfache Weise mit der viel zitierten „Siri"-Software der Firma Apple demonstrieren kann. Man muss z. B. einen Text in einen Browser tippen, die Übersetzungsfunktion einschalten und die Maschine bitten, den Text Siri vorzulesen. Was das Sprachverarbeitungssystem dazu bemerken wird, kann man ahnen. Was es nicht tun wird, kann man fest vorhersagen: etwas Vernünftiges.

Ein weiterer technikunabhängiger Nachteil der Spracherkennung besteht in ihrer Funktionsweise selbst. Im Prinzip sollte das Funktionieren einer Technik möglichst von der Umwelt unabhängig sein. Ebenso gilt, dass die Benutzung einer Technik die Umwelt möglichst wenig beeinflussen darf. Spracherkennung ist aber immer mehr oder weniger geräusch- und damit umweltabhängig. Zudem stört die Eingabe in ein solches System die Umgebung mehr oder weniger stark. Während z. B. eine Eingabe über die Tastatur sehr diskret ist, kann in einem Großraum jeder mithören, was man in das System spricht. Man stelle sich vor, in einer Sitzung würden alle Beteiligten ihre Notizen in ihre Maschinen sprechen!

In der Literatur wenig Beachtung finden übrigens die begrenzten Fähigkeiten des Menschen, über längere Zeit zu sprechen. Während es Mitarbeiter gibt, die mehrere Jahre täglich zehn Stunden und mehr tippen können, dürfte kaum jemand in der Lage sein, auch nur wenige Monate eine akustische Eingabe über diese Zeitspanne durchzuhalten. (Anmerkung: Das Projekt „CCall" über die Arbeit in Call-Centern enthält viele diesbezügliche Erkenntnisse.)

Ein Grund hierfür ist die Belastung der Ohren des Sprechers durch seine eigene Stimme, die etwa einem Lärmpegel von 105 dB(A) entspricht. Dem Gehör ist nämlich egal, wo der Lärm her kommt. Auch die Stimmbänder leiden stärker als üblich unter trockener Büroluft.

Was immer besser werden wird, aber nicht gleich

Abgesehen von technikunabhängigen Gesichtspunkten existieren technikbedingte Probleme, die zwar seit der Vorstellung der ersten Spracherkennungssysteme in den 1950er Jahren gewaltig zurückgedrängt worden sind, aber immer noch für mangelnde Akzeptanz sorgen. Dies liegt daran, dass linguistische Probleme nur mit einem riesigen Rechenaufwand, wenn überhaupt zu lösen sind. So benötigt Siri zum Funktionieren ein Rechenzentrum, obwohl die Maschine, auf der sie läuft, das iPhone 4S, eine größere Rechenpower hat als eben Mondlandefähre plus Bodenstation. Siri braucht ihr Houston. Was übrigens nicht nur ein betriebliches Datenschutzproblem darstellt.

Und das Sprachverarbeitungssystem muss sehr häufig nachfragen. Systeme, die nicht Siri heißen, arbeiten zuweilen mit „cloudbasierter Spracherkennung per Streaming-Technologie". Die wolkige Bezeichnung bedeutet nichts anderes, als dass das benutzte System häufig externe Hilfe benötigt.

Um sogenannte Homophone, also unterschiedlich geschriebene, aber vollkommen gleich klingende Wörter wie „fiel" und „viel", sicher zu verarbeiten, müsste die Maschine den Inhalt „verstehen". Da es wohl nur einen einzigen Computer gibt, der so etwas wie „Verstehen" hinreichend gut zuwege bringt, Watson, der bei Jeopardy, einem besonders kniffligen Quiz, einen Menschen geschlagen hat, wird man noch sehr lange warten müssen, bis das „Verstehen" funktioniert. Dabei hat Watson nicht etwa irgendeinen geschlagen, sondern einen Reihensieger.

Aber auch ohne solche kniffligen Sprachprobleme haben Computer Schwierigkeiten, mit der enormen Vielfalt an Ausdrucksformen zu eigentlich einer relativ geringen Zahl von Wörtern umzugehen. Die oben genannte Datenbank mit Millionen „Wörtern" enthält nicht etwa Millionen Exemplare aus dem Wortschatz einer Sprache, sondern z. B. ihre Deklinationen oder unterschiedliche Formen der Aussprache.

Da die Schwierigkeiten mit der Vielfalt der Sprache gleich zu Beginn aufgetreten sind, verfolgt man seit Langem eine zweigleisige Strategie. Spracherkennungssysteme werden sprecherabhängig bzw. sprecherunabhängig gestaltet. Sprecherabhängige Systeme benötigen ein mehr oder weniger langes Training beim gleichen Sprecher. Und manch technologiefreundlicher Benutzer hat, nachdem die Maschine nach einer halben Stunde immer noch nicht die Begrüßung „Guten Tag, meine Damen und Herren" beherrschte, Fortschritt hin, Fortschritt her, seine teure Anschaffung schnell entsorgt.

Ein wesentlicher Grund für eine derart schnelle Trennung liegt häufig darin, dass bei der Arbeit mit einem solchen System nicht nur die Eingabe über die Sprache erfolgt, sondern auch die Steuerung des Systems. Hierzu muss man eine mehr oder weniger komplexe Kommandosprache beherrschen und der Maschine stets korrekt mitteilen, ob das Gesprochene zum Inhalt gehört oder sich an die Maschine richtet. Keine geringe Anforderung an die Konzentrationsfähigkeit des Mitarbeiters.

Man wird sich nicht selten wundern, dass manches hochgelobte System nicht mit der Groß- und Kleinschreibung klar kommt. Das liegt nicht selten daran, dass Computer im Prinzip die Groß- und Kleinschreibung nie richtig gelernt haben. Hinzu kommt, dass sie bzw. ihre Betriebssysteme ausnahmslos aus den USA stammen, wo die Bedeutung der Groß- und Kleinschreibung viel geringer ist als in deutschsprachigen Ländern.

Die mangelhafte Fähigkeit der Systeme wird zum Teil durch branchenspezifische Wörterbücher ausgeglichen. Leider nur zum Teil, weil auch ein Jurist Begriffe aus dem Alltag oder aus der Sprache der Wirtschaft benutzen muss oder ein IT-Fachmann halt ein Spracherkennungssystem für Juristen entwickelt. Auch hier gilt, Angehörige von Berufen mit einer präzisen Fachsprache profitieren von solchen Einrichtungen wie Fachwörterbüchern. Beispielsweise kann man die Erstellung von Arztbriefen nahezu automatisieren. Der wesentliche Vorteil einer solchen Hilfe besteht nicht etwa darin, dass man die lästige Eingabe einfacher vornehmen kann, wofür man notfalls jemanden abstellen könnte. Vielmehr kann man einem „instant" erstellten Arztbrief gegenüber einem nachträglich geschriebenen einen Quantensprung in Sachen Qualität bescheinigen.

Qualität ist nicht alles, aber fast

Ein wesentliches Problem von Spracherkennungssystemen, wie übrigens auch von den bereits genannten OCR-Systemen, ist die „Qualität", womit die Erkennungsrate gemeint ist. Zwar liegen Zeiten, wo die Erkennungsrate 90 % betrug, lange hinter uns. Dennoch können auch solche Zahlen, die den Techniker in Ehrfurcht erstarren lassen, z. B. 99,9 %, einen Verwaltungsmenschen kalt lassen. Wenn zehn Seiten Text (30 000 Zeichen) zu 99,9 % korrekt ist, finden sich darin immer noch 30 falsche Zeichen, während eine Sekretärin bereits ein Zehntel davon als beschämend ansehen würde. Selbst ein Lohnschreibbüro müsste um den nächsten Auftrag bangen, wenn es eine solche Qualität ablieferte. Wenn es sich bei dem Text gar um Daten handelt, wird der Kaufmann auf solche Dienste dankend verzichten.

Reale Systeme erreichen eine Genauigkeit von etwa 99 %, das allerdings auch nur nach längerem Training. Das hieße beim Beispiel von zehn Seiten Text etwa 300 (!) Fehler. Alternativ kann man sich der Korrektur-Software bedienen, um die Fehler zu reduzieren. Eine zeitraubende Angelegenheit. Vor allem ist man nicht vor Überraschungen sicher, weil eine Korrektur-Software Fehler nicht immer generell reduziert, sondern zuweilen schwerer erkennbar macht.

Solchen Qualitätsproblemen ist zu verdanken, dass Spracherkennung und OCR, die Ende der 1980er Jahre als Pfeiler der kommenden IT-Systeme angesehen wurden, nie aus ihrem Nischendasein herauskommen konnten. Das liegt nicht zuletzt daran, dass die Bezeichnung „Spracherkennung" irreführend ist. Eine Maschine, die „A" von „B" unterscheiden kann, weiß noch lange nicht, was die beiden bedeuten.

Die obige Feststellung gilt nicht für solche Anwendungen, die einem ansonsten nicht erfüllbaren Ziel dienen wie z. B. bestimmte Systeme für Mediziner, die keine sonstige Eingabe machen können bzw. dürfen, so wenn sie mit ihren Händen etwas anderes machen müssen oder steril sind und deswegen keine Tastatur berühren dürfen.

Kommandosysteme

Sogenannte Kommandosysteme („Sprachbediensystem", „Voice Control System") sind zwar auch Spracherkennungssysteme, der Zweck der Erkennung liegt aber darin, den Inhalt zu „verstehen" und in Aktionen umzusetzen. Da Verstehen im Allgemeinen nicht zu den großen Stärken von Maschinen zählt, funktionieren sie durch eine Begrenzung des benutzten Wortschatzes. Dieser kann im Minimalfall nur aus dem bestehen, was auch ein Schalter „verstehen" kann: Ein – Aus.

Den Maximalfall, d. h. die komplexeste sicher erkennbare Kommandostruktur, bestimmt die technische Leistungsfähigkeit des Systems. Bei reinen Kommandosystemen übernimmt die Sprache die Rolle der Eingabeschnittstelle, für die auch eine Tastatur, ein Joystick oder ein Fußpedal denkbar wären.

Gut funktionierende Kommandosysteme existieren schon sehr lange. Bereits vor mehr als zwei Jahrzehnten konnte man z. B. Hochregallager für Menschen aus mehreren Ländern bedienbar machen, indem man akustische Befehle in der jeweiligen Sprache ermöglicht hatte. Heute kann man von Kaffeemaschinen bis Autos oder Häusern alle möglichen Alltagsgegenstände, aber auch komplexe Maschinerie über Kommandosysteme bedienen.

Ein Kommandosystem lässt sich auch mit einem Sprechererkennungssystem kombinieren und damit eine Anwendung schaffen, die nur einer einzigen Person gehorcht.

Kommandosysteme gehören zu den erfolgreichsten Anwendungen der Spracherkennung, was nicht verwunderlich ist, weil man bei deren Gestaltung bereits die Einschränkungen berücksichtigen kann, die bei einer Kommunikation mit gestörten bzw. funktionell eingeschränkten Mitteln vorkommen können. Im Prinzip bilden Kommandosprachen jeglicher Art eine Antwort auf beschränkte Kommunikationssysteme. Ihr wesentlicher Mangel, die Unnatürlichkeit der Kommunikation, ist zugleich ihre Stärke.

Sprachdialogsysteme

Sprachdialogsysteme bestehen aus Kombinationen von Sprach- bzw. Sprechererkennungssystemen, Kommandosystemen und Sprachausgabesystemen. Man kann sie seit einiger Zeit häufig erleben, wenn man größere Firmen anruft, die ihr Personal durch solche Einrichtungen gegen eingehende Anrufe abschirmen: „Wenn Sie … möchten, drücken Sie bitte die ‚1' Wenn Sie …". Wer kennt diese Ansagen nicht? Ihre geringe Beliebtheit bei den Anrufern mag zwar verschiedene Gründe haben, aber ihre gemeinsame Schwäche liegt in ihrer Funktionsweise: Geräusche sind eindimensional, ihre Dimension ist die Zeit. Während man bei einer optischen Anzeige (Bildschirm, Zeitung, Buch usw.) beliebig hin- und hergucken kann und eventuell rückwärts suchen, kann man bei einem akustischen Medium gleichzeitig nur eine Silbe hören und nur vorwärts. Deswegen können z. B. Ansagen wie die oben angeführte nur wenige Optionen enthalten, zudem muss die wichtigste Option zuerst kommen. Was, wenn man die Option verpasst hat?

Ob Sprachdialogsysteme eine große Zukunft vor oder eher hinter sich haben, lässt sich nicht genau sagen. Was sich mit Bestimmtheit sagen lässt, ist, dass sie nicht barrierefrei sind. Während man bei den ebenso nicht barrierefreien optischen Darstellungen (Text) recht gut in gesprochene Form umsetzen kann, lässt sich eine akustische Information, wie oben erklärt, nur mühsam und unvollkommen in eine optische umwandeln.

Aussichten

Spracherkennung sowie die weiteren Anwendungen, die hier betrachtet wurden, leiden an drei wesentlichen Problemen, die eng mit den Eigenschaften der gesprochenen Sprache zusammen hängen. Diese sind nicht oder schwer behebbar. Die natürlich gesprochene Sprache ist zwar ein wunderbares Instrument für den menschlichen Dialog, lässt sich aber für andere Zwecke relativ mühsam missbrauchen. Daher wird ihre Bedeutung trotz technischen Fortschritts allgemein gering bleiben. Das will aber nicht heißen, dass solchen Anwendungen ein Erfolg auf vielen Spezialgebieten versagt bleiben wird. In manchen Fällen werden sie oft die bestmögliche Anwendung bleiben oder werden.

Autor

Dr.-Ing. Ahmet E. Çakir, Leitung des Ergonomic Instituts für Arbeits- und Sozialforschung, Beratung von Anwendern und Herstellern in allen Fragen der ergonomischen Gestaltung von Bildschirmarbeit; ahmet.cakir@ergonomic.de, www.ergonomic.de