



# Reden ist Silber, „verstehen“ ist Gold

Voice Portale auf dem Weg zum natürlichen Dialog zwischen Mensch und Maschine.

Von Detlev Artelt

Die Nutzung natürlicher Spracherkennung am Telefon kann als ausgereift bezeichnet werden und eröffnet mittlerweile eine Vielzahl an Möglichkeiten zur Kosteneinsparung im Unternehmen. In wirtschaftlich schwierigen Zeiten wie heute sind Einsparungen an jeder Stelle mehr als erwünscht, weshalb Voice Portale (Definition Voice Portale siehe Kasten auf S. 10) auch immer mehr Zuspruch finden. So können durch eine automatische Rufnummernauskunft immense Call Center-Kosten reduziert oder durch ein Firmen-Voice Portal Kunden rund um die Uhr mit Informationen per Self-Service versorgt werden. Und all dies ohne die ansonsten so

hohen Personalkosten. Ein Voice Portal hat laut der Studie „Ist Deutschland sprachlos?“ (siehe Computer TELEPHONY Voice Business 2003) der Unternehmensberater Strateco und MIND einen durchschnittlichen ROI von nur sieben Monaten.

In Deutschland werden, im Vergleich zu den USA oder anderen Ländern, solche Neuerungen immer erst nach einigen Jahren von der Masse akzeptiert, aber ein Trend zu Voice Portalen ist bereits deutlich erkennbar. „Es ist eine der zentralen Erkenntnisse unserer Untersuchungen, dass die vermuteten Akzeptanzbarrieren nicht primär in der mangelnden Nutzerakzeptanz zu suchen sind,

sondern auf Widerstände bei Entscheidungsträgern zurückzuführen sind", meint Jens Klemann, Geschäftsführer der Strateco GmbH und Mitinitiator der Initiative Voice Business. Die allseits bekannten Kinoauskunftssysteme oder oftmals schlecht konzipierte IVR-Systeme (Interactive Voice Response) mit Einzelworterkennung gelten als veraltet und nicht mehr repräsentativ, sind aber immer noch in den Köpfen der Entscheider, wenn es um die Entscheidung für Voice Portale geht. Ähnliche Entwicklungen lassen sich bei der Einführung vieler anderer Techniken beobachten. Beispielsweise wollte vor zehn Jahren niemand einen Anrufbeantworter verwenden, heute nutzt jeder mehrere davon, gleich ob sich dieser Anrufbeantworter zu Hause oder am Mobiltelefon im Netz des Providers befindet. Auch zur Einführung von Anrufbeantwortern konnte man durchaus viele Geräte finden, die die Nutzer eher abschreckten. Doch hier wie dort gilt: Ist der Nutzen erst mal auf breiter Fläche akzeptiert, kommt auch der Erfolg der Technik.

### Markt der Zukunft

Die Zeichen für den Einsatz von Sprachlösungen stehen sehr gut, denn das Marktvolumen wächst augenblicklich und soll laut Gartner Dataquest in den Jahren bis 2008 ein jährliches Wachstum von etwa 30 Prozent erreichen. Dies freut die Lieferanten von sogenannten ASR-Produkten (Automatic Speech Recognition), obwohl sich der Preis pro Einheit deutlich nach unten bewegen soll. Der Erfolg hat aber wie immer viele Väter: Einerseits hat die Vielzahl der installierten Systeme mit Spracherkennung den Beweis erbracht, dass Spracherkennung funktioniert und dem Kunden einen Nutzen erbringt. Zudem werden die ASR-Basistechnologien immer perfekter und die internet-basierenden Standards wie VoIP und VoiceXML erfreuen sich zunehmender Beliebtheit. Viele ehemals noch proprietäre Schnittstellen werden durch neue Standards ersetzt. Im Bereich der Spracherkennung ist dies das Media Resource Control Protocol, kurz MRCP genannt. Das MRCP wurde bereits 2001 von ScanSoft, Cisco and Nuance definiert, Produkte auf dieser Basis sind mittlerweile vorhanden. Durch solche Standards werden die einzelnen Sprachprodukte auch immer austauschbarer, denn nun lässt sich ein Spracherkennung von Lieferant A durch einen Spracherkennung von Lieferant B ersetzen, wobei die Praxis hierbei sicherlich noch den einen oder anderen Stolperstein mit sich bringt.

Der Weg, den Anbieter im Bereich Sprachverarbeitende Systeme bis heute zurückgelegt haben, war besonders in den vergangenen fünf Jahren mit dem Niedergang der TK-Industrie schwer. Die Folge war eine Konsolidierung unter den Spracherkennungsanbietern. Namen wie Lernout & Hauspie oder Philips verschwanden, die großen Anbieter stehen fest. Die Tabelle auf Seite 10 zeigt den Markt in EMEA und weist laut Gartner die Firma Scansoft als Marktführer aus. Andere Anbieter teilen sich die Nischen, oder liefern beispielsweise Erkennen, die besonders gut für bestimmte Sprachen geeignet sind. Mit der Gründung der Acapela Group, der unter anderem Hersteller wie Elan beitraten, sollte

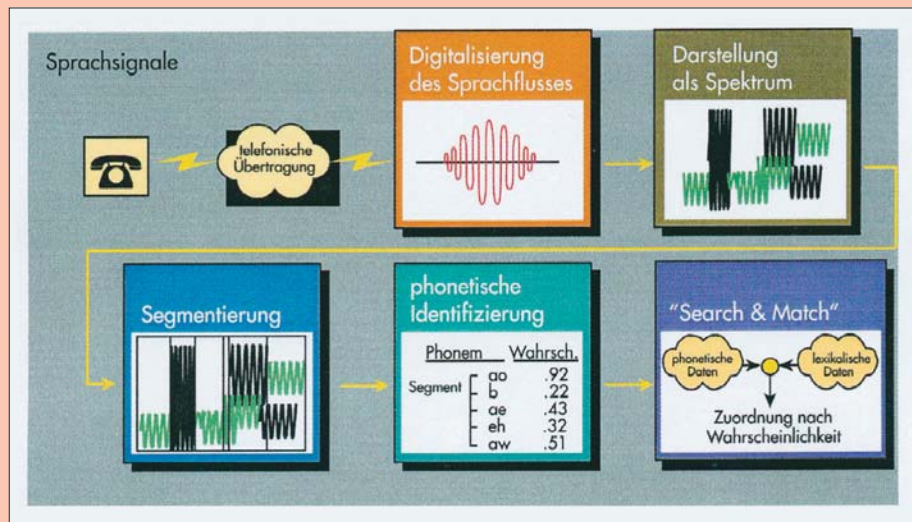
im Markt ein Gegenpol zur Marktmacht Scansofts aufgebaut werden, die Ergebnisse dieser Bemühungen stehen allerdings noch aus.

„Wir wollen dem Kunden neben der Basistechnologie mit Teilapplikationen helfen, schneller und kosteneffektiver Voice Portale entwickeln zu können“, definiert Reimund Schmalz von Scansoft die Zielsetzung des Marktführers und führt aus: „Unsere Kunden müssen nicht immer wieder das Rad neu erfinden lassen, sondern bekommen von uns gleich einen Sub- oder Teildialog, ein sogenanntes Speechpack, beispielsweise für eine Telefonnummernauskunft, mitgeliefert.“ Scansoft sichert seine Marktposition durch

## So funktioniert Spracherkennung

Spracherkennungsanwendungen funktionieren im Grunde ähnlich wie der menschliche Hörvorgang, bei dem das Ohr Laute in Form von Schwingungen aufnimmt und das Gehirn die Signale decodiert und deren Bedeutung bestimmt. Spricht ein Anrufer mit einem Dialogsystem, fängt das Spracherkennungssystem zunächst seine Äußerung auf, digitalisiert sie und stellt sie als Spektrum dar. Dabei wird die Äußerung des Anrufers in einzelne phonetische Segmente, also die kleinsten bedeutungsdifferenzierenden sprachlichen Einheiten, zerlegt. Im nächsten Schritt müssen die einzelnen phonetischen Segmente identifiziert werden. Diese äußerst schwierige Aufgabe bewältigt das System, indem es die analysierten Segmente mit einer Vielzahl von komplexen statistischen Modellen vergleicht und die Phoneme mit der größten Übereinstimmung auswählt. Schließlich

muss den ermittelten Lauten und den aus ihnen ableitbaren potenziellen Äußerungen die im jeweiligen Zusammenhang sinnvollste Bedeutung zugeordnet werden. Diese Aufgabe löst das System mit Hilfe von hinterlegten Grammatiken, Wörterbüchern und Sprachmodellen. Die Sprachanwendung entscheidet sich für die jeweils wahrscheinlichste Übersetzung und erstellt hierfür einen so genannten „confidence score“, der angibt, wie sicher das System ist, die Äußerung korrekt verstanden zu haben. Ein hoher „confidence score“ führt dazu, dass das System das Gespräch mit dem Anrufer auf der Basis der ermittelten Bedeutung weiterführt. Ein eher niedriger „confidence score“ wird das System dazu veranlassen, nachzufragen, ob es eine Aussage richtig verstanden hat. Auf diese Weise wird ein durchgehender und sinnvoller Dialog zwischen Mensch und Maschine möglich.



Der Spracherkennungsprozess in schematischer Darstellung

Produkte wie multilinguale Spracherkennung, Text-to-Speech-Systeme und Stimmerkennungstechnologie, die ebenfalls zum Portfolio zählen.

## Sprache aus dem Computer

Ein wesentlicher Bestandteil von Voice Portalen ist die Text-to-Speech-Komponente. Auch hier ist die Entwicklung in den letzten Jahren weit vorangeschritten und eröffnet neue Per-

ASR-Topflieferanten in EMEA in 2003		
Hersteller	Ports	Marktanteil
ScanSoft	11.314	40,80 %
Telisma	5.479	19,80 %
Nuance Communications	4.245	15,30 %
Loquendo	2.47	8,90 %
IBM	1.525	5,50 %
Fluency Voice Technology	1.143	4,10 %
Sakrament	300	1,10 %
andere	1.263	4,60 %
Total	27.739	100 %

spektiven: Jedes Unternehmen besitzt eine Vielzahl von Datenbanken mit Produktinformationen, die nun vom Computer wie von einer menschlichen Stimme gesprochen werden können. Die Technik findet immer mehr Anklang und das nicht nur im Bereich der Voice Portale. Wie sehr sich die Qualität der Systeme entwickelt hat, lässt sich durch Tests leicht feststellen: Nur wenn man ganz genau hinhört, lassen sich bei aktuellen Lösungen noch so etwas wie leichte Gluckser oder falsche Betonungen hören. Die Zeit, in der TTS-Systeme sofort anhand der gebrochenen Betonung in einem Satz erkannt werden konnten, ist vorbei: Selbst audiophil-professionelle Rundfunkstationen setzen mittlerweile Text-to-Speech-Systeme ein, um Informationen auch per Telefon an den Kunden zu bringen.

## Das eigene Voice Portal

Unternehmen, die Sprachtechnik anwenden wollen, sind gut damit beraten, sich einen aktuellen Marktüberblick zu verschaffen, denn das Angebot der notwendigen Komponenten ist stark gewachsen. Zu den „Zutaten“ einer Sprachlösung zählen einzelne Basistechnologien wie Spracherkennung oder Text-to-Speech-Komponenten, Applikationsgeneratoren wie ISDN-Karten-gestützte automatische Telefonzentralen, bis hin zu als Dienstleistung gehosteten Sprachdialogplattformen. Mittlerweile sollte für nahezu jede

Anforderung im Unternehmen eine passende Lösung zu finden sein. Ist die Zielanwendung sehr komplex, empfiehlt sich ein Blick auf die am Markt verfügbaren Applikationsgeneratoren, die nahezu jedem Problem gewachsen sind. Diese wiederum unterstützen eine Vielzahl an Spracherkennern und ISDN-Karten, so dass sich hieraus das Wunschsystem ableiten lässt. Voraussetzung ist hier allerdings, dass auch das entsprechende personelle Knowhow für die Programmierung von Applikationsgeneratoren im Unternehmen vorhanden ist.

Alternativ stehen Dienstleister zur Verfügung, die die Zielapplikation entwerfen und wie eine Website bei einem Voice Service Provider hosten können. In einem solchen Fall wird die Applikation nur einmalig erstellt und dann auf den bestehenden Voice Portalen des Dienstleisters gehostet. Das erspart die leider immer noch sehr hohen Investitionskosten in die ISDN-Hardware und Sprachsoftware. Generell empfiehlt sich

hier ein Test dazu, ob die Zielkunden überhaupt ein Voice Portal akzeptieren oder lieber weiterhin mit dem persönlichen Kontakt

sprechen wollen. Bei einem solchen Test lassen sich auch die potenziell vorhandenen Integrationsprobleme erkennen und gezielt umschiffen.

## Immer den Überblick behalten

Unternehmen, die sich bisher kaum oder nur wenig mit Sprachtechnik beschäftigt haben, sei der Rat des Fachmanns dringend empfohlen. Durch die falsche Kombination von Technologie und Dienstleistung kann sehr schnell ein teurer Fehlgriff entstehen. Mit der Idee für ein eigenes Voice Portal und einem passenden Berater kann sich jedes Unternehmen vor den typischen Anfangsfehlern schützen. Zwar kann man auch einen seiner eigenen Mitarbeiter daran setzen, mit einem Applikationsgenerator die individuelle Dialogstruktur einer Telefonzentrale selbst zu entwickeln. Aber: Sprachtechnik ist sehr komplex und erfordert ein hohes Knowhow und entsprechenden Zeitaufwand bei der Programmierung. Fertiglösungen lassen sich schnell einsetzen, bieten aber in der Regel eine sehr geringe Implementationstiefe und Flexibilität. Sie können beispielsweise bei Telefonanfragen nicht auf hinterlegte Inhalte von Unternehmensdatenbanken zugreifen,

## Was ist ein Voice Portal?

Seit Jahren sind uns die unterschiedlichsten Systeme bekannt, die am Telefon automatisch einen Dialog ermöglichen. Sei es eine Mailbox oder ein Bankingsystem, welches den Kontostand an das Telefon bringt. Hierbei wird in der Regel eine Frage gestellt, die vom Benutzer mit der Eingabe einer oder mehrerer Ziffern per Telefontastatur beantwortet wird. Diese Systeme werden IVR (Interactive Voice Response) genannt, da der Benutzer interaktiv den Ablauf des Dialogs bestimmen kann. Die Erkennung der Antwort wird über die Töne (DTMF) des Telefons in den Computer „gepiepst“. Bei vielen Auswahlmöglichkeiten stößt ein IVR-System schnell an seine Grenzen, denn wenn ein komplexer Abfragebaum aufgestellt werden muss, wird der Dialog sehr lang und unübersichtlich.

Bei einem Voice Portal ist das anders, denn Voice Portale arbeiten im Gegensatz zu den IVR-Systemen mit Spracherkennung und mit Text-to-Speech, um den Benutzer zu führen. Dies ermöglicht dem Benutzer ein viel größeres Antwortspektrum, da er in normalen Worten oder ganzen Sätzen antworten kann. Voice Portale werden auch

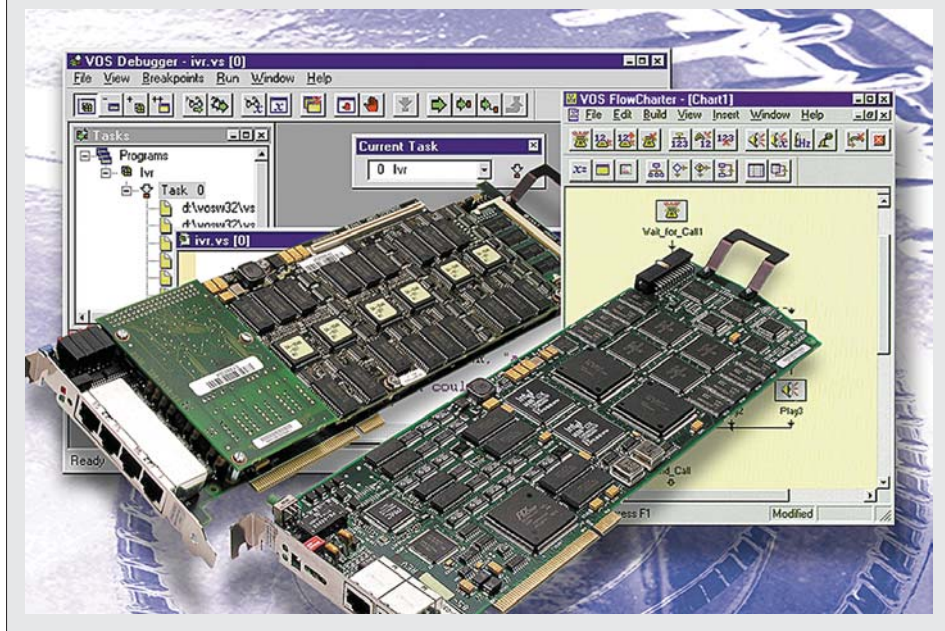
gern als die Nachfolger der IVR-Systeme bezeichnet. Erkennt das Voice Portal die Worte oder den gesprochenen Satz des Benutzers genau genug, so wird der Dialog mit einer Antwort fortgesetzt, die oftmals aus einer Datenbank abgefragt wird und einfach per Text-to-Speech vom Computer vorgelesen wird.

Aus technischer Sicht besteht ein Voice Portal aus mehreren Komponenten. Dies sind der Applikationsserver, der Spracherkennung, das Text-to-Speech-System und eine Dialogsoftware. Die Softwareplattform auf Basis einer VXML-Plattform oder eines Applikationsgenerators verbindet nun alle diese Technologien und ermöglicht die Programmierung einer individuellen Lösung. Über ISDN-Karten oder Voice-over-IP-Gateways wird das Portal mit dem Telefonnetz verbunden. Bleibt noch zu erwähnen, dass zur Zeit die Auswahl der Plattform Software manchmal eher einer Glaubensfrage gleichkommt. Hier stehen die erwachsenen IVR-Plattformen den reinen VoiceXML Plattformen gegenüber, wobei die IVR-Plattformen oftmals ebenfalls alle VXML-Funktionen zur Verfügung stellen.

## Nutzung von Sprachlösungen

Aktuelle Sprachlösungen werden aus mehreren Komponenten aufgebaut (siehe Kasten links unten), die untereinander austauschbar sein sollten. Das gilt nicht nur für die Hardware, sondern auch für die Software. Durch die technischen Fortschritte in den letzten Jahren ist es für Unternehmen allerdings nicht unbedingt mehr notwendig, die für eine Sprachlösung benötigte Hard- und Software selbst zu kaufen. Hostinganbieter wie die Deutsche Telekom oder Speeachart erlauben es Unternehmen, Applikationen

zu nutzen, für die keine dedizierte im Unternehmen bereit stehen muss. In jedem Fall empfiehlt sich vor Einsatz einer Sprachlösung das Hinzuziehen eines Beraters, sofern das für die Programmierung der Applikation notwendige Knowhow nicht inhouse zur Verfügung steht. So kann die Applikation von Profis entworfen, und von einem Provider auf dessen Plattform gehostet werden. In der Regel empfiehlt es sich ebenfalls, einen kompletten Piloten vorab erstellen und testen zu lassen.



weil entsprechende Schnittstellen fehlen. Wer sich tiefer in Sprachtechnik einarbeiten möchte, der sollte sich intensiver mit Applikationsgeneratoren und VXML-Plattformen beschäftigen und ebenfalls nach einem Linguisten bzw. einem geeigneten Berater im Bereich Voice Design Ausschau halten.

### Ausblick

Voice Portale sind ernstzunehmende Businessanwendungen geworden, die mit einem ROI von sieben Monaten nicht nur große Unternehmen interessieren sollten. Durch die vermehrt aufkommenden Voice Portal Hosting Provider lassen sich die ausgereiften Basistechnologien kostengünstig für Standardanwendungen nutzen. Bei komplexeren Anwendungen empfiehlt sich aber nach wie vor das eigene System, um flexible Anpassungen an eigene Datenbanken zu realisieren. Wichtig ist, dass der Dialog aus Nutzersicht erstellt und im Ablauf möglichst einfach gehalten wird. Auch hier empfiehlt sich

das Fachwissen eines kompetenten Beraters (siehe Marktübersicht Voice-Dienstleister ab Seite 54).

## Kontakt



Detlev Artelt ist Geschäftsführer der aixvox GmbH in Aachen.

Aixvox GmbH  
Kontakt: (0241) 4133-0 da@aixvox.com  
Internet: www.aixvox.com