**Sprach-Portale ermöglichen schon jetzt einen besseren Kundenservice beziehungsweise den Zugriff auf Informationen per Telefon. Multimodale Lösungen, also die Kombination von Sprache, Text und Grafiken für die Ein- und Ausgabe, sollen die Akzeptanz und den Nutzen noch erhöhen.**



Bild: IBM

# Die Zukunft der Voice-Portale

Von Christoph Kartes

Viele mobile Mitarbeiter verbringen eine Menge Zeit in Autos oder an Plätzen, wo ein Telefon die einzige Kommunikationsmöglichkeit ist. So erlauben zwar immer mehr neue mobile Geräte Remote Access, jedoch frustrieren die kleinen Displays und schwierigen Eingabemöglichkeiten oft den Benutzer. In diesen Fällen hilft der Zugang per Sprache und auch die Sprachausgabe. So ist es etwa viel einfacher, einen Namen aus einem Adressbuch zu sagen als ihn per Eingabe zu suchen.

Bisher war das größte Hemmnis für Sprach-Applikationen die unnatürliche Computer-generierte Stimme. Doch technische Fortschritte lassen die PC-Stimmen mittlerweile fast menschlich klingen. Auch die Genauigkeit der Spracherkennung hat sich weiterentwickelt, sodass auch dies keinen Hinderungsgrund mehr darstellt. Zu guter Letzt wurden offene Standards veröffentlicht, die es Unternehmen erlauben, Voice-Applikationen mit geringen Aufwand in bestehende Lösungen zu integrieren.

## Standards und Entwicklungen

Sprach-Applikationen lassen sich derzeit mit Hilfe von zwei Programmiersprachen erstellen: VoiceXML und Salt (Speech Application Language Tags). Der offizielle Standard ist VoiceXML, das mittlerweile in der Version 2 vorliegt. Das W3C (World Wide Web Consortium) arbeitet derzeit an der Version 3, die Anfang nächsten Jahres als Draft erscheinen soll. Diese enthält dann auch weitere Ansätze für Multimodalität. Die Vorteile von VoiceXML liegen in der Standardisierung und der einfachen Programmierung. Theoretisch können Entwickler von Webseiten, die HTML, XML und Java beherrschen, auch Sprach-Lösungen erstellen. Außerdem sind Standard-Systeme kompatibel zueinander.

Genau da liegt aber das Problem. Auf Grund des vergleichsweise geringen Funktionsumfangs von VoiceXML hat so gut wie jeder Hersteller proprietäre Erweiterungen vorgenommen, welche die Portabilität von Applikationen behindern. So hat etwa Genesys in sein Produkt TXML integriert (TeleraXML), wodurch unter anderem erweiterte Anrufkontrollmöglichkeiten bereit gestellt werden. Doch unterstützt Genesys natürlich auch Standards wie etwa CCXML (Call Control Extensible Markup Language).

Darüber hinaus liegt in der Dialog-Entwicklung die größte Herausforderung: Zwar können Web-Programmierer Voice-Applikationen erstellen, doch bedarf es laut dem Anbieter von Kommunikationslösungen Sikom anderer Qualifikationen, als gute Webseiten erstellen zu können. Eine umfassende Erfahrung im Umgang mit gesprochener Sprache ist am Wichtigsten. Dabei kann bei der Implementierung davon ausgegangen werden, dass das Dialog-Design genauso viel Zeit in Anspruch nimmt, wie die Appliktions-Integration selbst.

Neben VoiceXML ist auch Salt im Gespräch, wenn es um Sprach-Applikationen geht. Doch Salt ist nur als Draft vorhanden und liegt beim W3C zur Verabschiedung vor. Genauso wie X+V (XHTML + VoiceXML), das ebenfalls dem W3C vorliegt. X+V wurde außerdem beim VoiceXML-Gremium eingereicht , damit es in Version 3 von VoiceXML Berücksichtigung findet.

Zwar unterscheiden sich alle drei Möglichkeiten in vielen Punkten, doch sie „verwenden dieselben Grundlagen wie zum Beispiel SRGS (Speech Recognition Grammar Specification) für Grammatiken, CCXML für Call Control und SSML (Speech Synthesis Markup Language) für Sprachsynthese-Formatierung“, so Tiemo Winterkamp, Leiter Product Management & Strategy von Voiceobjects. „Die Diskussion , welcher Standard nun eingesetzt werden soll, ist demnach nur oberflächlich.“

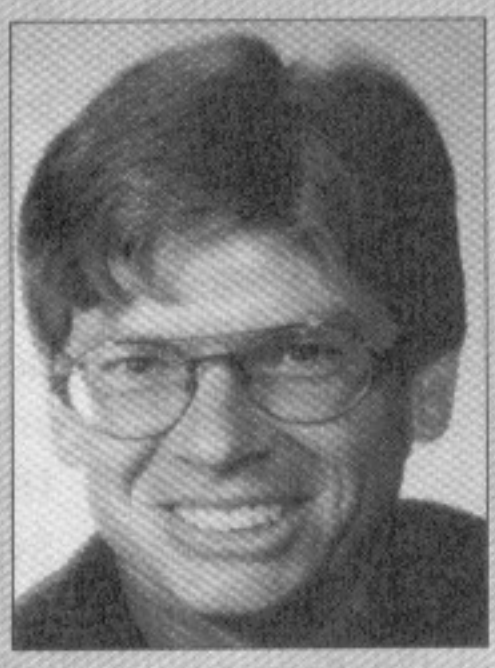
Welcher dieser Standards sich durchsetzt, ist noch nicht klar. Clarity glaubt, dass VoiceXML die beste Alternative für Multimodalität ist, jedoch nur in Verbindung mit EMMA (Multimodal Annotation Language, www.w3.ord/TR/ EMMAreqs/). „Salt ist nicht schlecht für einfache Lösungen, aber nicht leistungsfähig genug für wirklich anspruchsvolle multimodale Anwendungen“, so Marcus Munk, CTO der Clarity AG. Die W3C Multimodal Interaction Group etwa arbeitet an einem Grundgerüst. Hersteller wie Scansoft wirken aktiv daran mit und wollen die Ergebnisse voll unterstützen.

Bereits heute gibt es einige multimodale Anwendungen auf dem Markt. Dabei handelt es sich jedoch um Lösungen, die immer nur einen Kanal unterstützen – wenn beispielsweise per Sprache Infor-

## Standard-basierte mobile Internet-Anwendungen

Wolfgang Karbstein, Sales Manager Pervasive Computing & Voice Systems bei IBM: „Die multimodale Technologie wird die Art der Nutzung mobiler Geräte verändern wie es die Einführung der GUI (Graphical User Interface) auf dem PC getan hat. Als nur Texteingaben (zum Beispiel über DOS) möglich waren, nutzten nur wenige Anwender den PC, da der Umgang schwierig und die Eingabe von Informationen mühsam war. Seit Einführung der GUI nutzen Millionen den PC für mehr und mehr Anwendungen. Wir nehmen an, dass über die nächsten Monate und Jahre die meisten mobilen, Web-basierten Anwendungen für die Nutzung von multimodalen Geräten und Abläufen designed werden. Damit wird der momentan noch mühsame Aufwand, Daten innerhalb eines 'Modus' ein- und auszugeben, deutlich vereinfacht. Dies ist der Hintergrund für die Marktprojektionen für multimodale Technik und der Grundstein für ein deutlich verändertes und erweitertes Nutzungsspektrum mobiler Geräte.

Bild: IBM



**Wolfgang Karbstein ist Sales Manager Pervasive Computing & Voice Systems bei IBM**

IBM beteiligt sich intensiv an der Entwicklung Standard-basierter Lösungen. In der Multimodal User Group des W3C unterstützen wir als treibende Kraft den X+V-Ansatz, um einen Standard für multimodale Interaktionen zu kreieren. Dieser kombiniert VoiceXML, welches das Navigieren durch Webseiten mit Hilfe von Sprachkommandos unterstützt, mit XHTML, einer Markup Language zur Erstellung von Webseiten. Das Ziel ist, den Entwicklern eine Erweiterung von schon vertrauten Umgebungen anzubieten, die ihnen erlaubt, mobile Internetapplikationen standardbasiert zu erstellen."

## Neue Qualität von Anwendungen

Jörg Emonts, Geschäftsführer bei Telenet: „In der Kommunikation zwischen Mensch und Maschine werden multimodale Anwendungen eine neue Qualität von Anwendungen ermöglichen. So ist zum Beispiel die Steuerung eines Navigationssystems im Fahrzeug per Sprache wesentlich sinnvoller als die heute übliche Eingabe von Städten und Straßennamen über winzige Tasten.

Bild: Telenet



**Jörg Emonts ist Geschäftsführer bei Telenet**

Je nach Anwendung und Einsatzort können zukünftige Endgeräte wahlweise über die Sprache, Tastatur oder per Stift (PDA) bedient werden. So kann in der Navigation eine Stadt per Spracheingabe gewählt werden, ein Stadtplan erscheint als Grafik auf dem Display und der Anwender zeigt per Stifteingabe auf den Stadtteil, der ihn näher interessiert. Wünscht er eine Zielführung zu diesem Punkt, so wird er vom Endgerät mit Hinweisen auf dem Display und über Sprachanweisungen dort hin geleitet.

Gerade beim mobilen Einsatz im Fahrzeug ermöglicht die Sprache eine Nutzung von Diensten ohne Einschränkung der Verkehrssicherheit. Letztendlich wird die multimodale Kombination bei der Ein- und Ausgabe von Sprache, Text, Grafiken oder auch Videos auf einem Endgerät eine deutliche Verbesserung im Bedienkomfort bringen."



Bild: IBM

**Mobile Außendienstmitarbeiter sind häufig im Auto unterwegs. Die einzige Möglichkeit, während dieser Zeit auf Unternehmens-Informationen zuzugreifen, ist das Mobiltelefon**

mationen abgefragt werden, ist der Voice-Kanal geöffnet, erfolgt die Ausgabe auf einen PDA, wird der Voice-Kanal geschlossen und ein Datenkanal bereit gestellt. Der Grund dafür ist, dass heutige GSM-Netze (auch GPRS) nicht in der Lage sind, beide Möglichkeiten gleichzeitig zu übertragen. Doch auch mit der heutigen Technik lassen sich einige interessante Applikationen bereitstellen. So bietet etwa Genesys mit ihrem Voice-Portal eine Möglichkeit an, Informationen per E-Mail, Fax oder SMS über ein Gateway auszugeben. Solche Lösungen haben auch IBM auf Grundlage des Websphere Portal Server, SVOX oder Intervoice im Angebot, wobei die Ausgabe von zum Beispiel Tabellen oder größeren Auswahl-Listen (etwa Abflugzeiten, Zugverbindungen) im Vordergrund steht. Ähnlich geht Nuance an diese Thematik, wobei der Hersteller zusammen mit Partnern multimodale Applikationen anbietet. Darüber hinaus arbeitet auch Scansoft an solchen Lösungen. „Es gibt einige multimodale Applikationen im Pilotstatus, aber noch keine wurde kommerziell eingeführt. Scansoft will ein solches System mit Fokus Location Based Services im ersten Halbjahr 2004 auf den Markt bringen", so Hugo Barra, Product Manager, Multimodal Technologies von Scansoft.

## Interaktion mit Applikationen

Echte, simultane multimodale Anwendungen werden jedoch erst mit der Einführung von UMTS entstehen. Denn nur das UMTS-Netz ist in der Lage, Sprache und Daten gleichzeitig zu übertragen. „Simultane Multimodalität ist heute sicherlich noch Stand der Erforschung, denn massenfähige Produkte auf Server- als auch auf Endgeräteseite, sowie prägnante Anwendungsfälle sind noch nicht vorhanden", meint Winterkamp. Doch auch nach der Einführung von UMTS wird es noch einige Zeit dauern, bis simultane Sprach-Daten-Applikationen nutzbar sind.

Der Grund: Die Einführung solcher Dienste lohnt sich erst, wenn eine genügend große Abdeckung mit UMTS-Endgeräten vorliegt. Experten gehen davon aus, dass in etwa zwei Jahren mit einer breiten Einführung von Voice-Diensten zu rechnen ist.

Bis dahin müssen die Hersteller auch noch einige Hindernisse überwinden. So arbeiten bereits einige Unternehmen wie IBM, Microsoft oder Voicegenie an der Synchronisation von Sprache und Daten. Das gemischte Aussenden funktioniert schon, doch das Einsammeln von zum Beispiel einem Mix aus Daten- und Spracheingabe stellt die Entwickler vor Probleme. Doch die Hersteller sind zuversichtlich, bald erste funktionierende Prototypen zu präsentieren.

Auf der Anwenderseite stellt sich jedoch die Frage, wie die Ressourcenverwaltung funktioniert. Der Grund: Die Spracherkennung muss in der Regel lokal auf dem Endgerät laufen. In diesem Punkt muss eine Lösung gefunden werden, wer diese Applikationen verwaltet und wie die Administration ablaufen soll. Microsoft arbeitet derzeit etwa an einem Plug-in für den Browser, IBM wird wohl einen ähnlichen Weg gehen und ein Java-Applet veröffentlichen.

## Migration auf Multimodalität

In der Praxis sind normalerweise Applikationen vorhanden, auf die die Sprach- beziehungsweise multimodale Applikation aufgesetzt werden muss. Dabei spielt es keine Rolle, ob zuerst eine Voice-Lösung oder sonstige Datendienst vorhanden sind. Durch die Verwendung offener Standards lassen sich die Applikationen ohne großen Integrationsaufwand auf vorhandene Lösungen aufsetzen. Jedoch müssen Unternehmen beachten, dass Sprache eine andere Menüführung, wie zum Beispiel eine Internetseite, haben muss. Soll also eine vorhandene Web-Applikation per Sprache zugänglich werden, so müssen in der Webseite Dialoge hinterlegt sein, welche die Mausführung ersetzen. Dies bedeutet jedoch einen enormen Aufwand. Einfacher geht es, wenn ein eigener Service für den Sprachzugriff erstellt wird. Zum Beispiel wäre es denkbar, die Wissensdatenbank eines Unternehmens per Telefon auch Außendienstmitarbeitern oder Kunden zugänglich zu machen. Doch kann ein Anrufer mit einer umfangreichen Suchfunktion nicht viel Anfangen, besser wäre es, wenn er einfach eine Frage stellen kann. Das System nimmt diese Eingabe als Grundlage für die Suche und gibt möglicherweise einige Lösungen zur Auswahl an. Als Faustformel gilt: Je komplizierter der Aufbau einer Webseite, desto schwieriger die Umsetzung. *(CK)*

## Stimmen aus der Branche



Bild: Clarity

**„Die Frage ist: Wie kann ich Datendienste durch Sprachanwendungen bereichern?", so Marcus Munk, CTO von Clarity. „Da mobile Endgeräte immer kleiner, deren Bildschirme aber größer werden, wird die Bedienung der Geräte durch Sprache immer wichtiger. Das heißt beispielsweise, Dienste wie MMS bekommen eine sprachliche Benutzeroberfläche"**



Bild: Intervoice

**Cornelia Schwegler, Geschäftsführerin von Intervoice Deutschland: „Multimodale Anwendungen haben eine viel versprechende Zukunft, da sie eine hohe Bedienbarkeit einerseits und eine Produktivitätssteigerung andererseits versprechen. Entscheidend für den Erfolg solcher Anwendungen ist aber das grundlegend neu zu definierende Verhalten der Anwender. Der Nutzer muss immer im Mittelpunkt bleiben, und abhängig vom jeweiligen Equipment, seine Ansprüche flexibel bedient werden. Gerade auf diesem Gebiet haben wir in den letzten Jahren enorm viel dazu gelernt"**

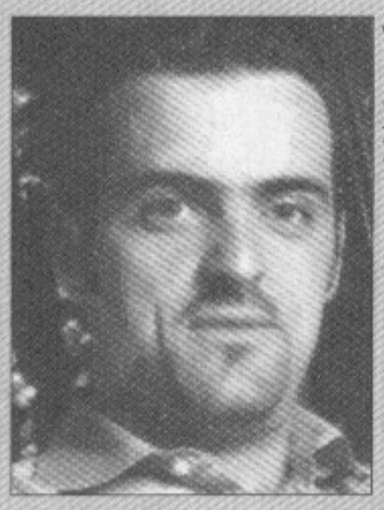


Bild: Scansoft

**„Benutzer werden bei multimodalen Anwendungen von einer schnelleren, benutzerfreundlichen und bequemeren Schnittstelle profitieren, die ein viel effizienteres Informationsangebot bietet, als das reine Sprach- oder Daten-Interfaces erzielen kann", glaubt Hugo Barra, Product Manager, Multimodal Technologies bei Scansoft**



Bild: Voiceobjects

**Tiemo Winterkamp, Leiter Product Management & Strategy von Voiceobjects: „Multimodalität wird in Zukunft ein wesentliches Interaktionsmerkmal sein, das vor allem dort zum Einsatz kommt, wo andere Interaktionen schwierig sind, wie etwa das Schreiben auf einem Mobiltelefon. Voiceobjects ist auch heute schon in der Lage, erste Anforderungen hierfür zu erfüllen, richtet sich aber nach den sich abzeichnenden Standards"**