Wertsteigerung

Video, Text und Grafik werten Sprachportale auf

Michael Codini

Bereits in vier Jahren wird jeder zweite Handybesitzer mobile Internetdienste nutzen, schätzt Forrester Research – und belegt damit das Interesse an multimedialen Services sowie die neuen Chancen für Sprachportale. Diese können mittlerweile deutlich mehr als nur Sprachein- und -ausgaben verstehen. Netzbetreibern, aber auch Inhalteanbietern und Unternehmen bieten sich damit neue Möglichkeiten, Umsätze zu generieren und die Kundenbindung zu erhöhen.

Michael Codini ist Chief Technology Officer bei der VoiceObjects AG in Bergisch Gladbach

Multimediale und multimodale 3G-Services wurden möglich durch die faktische Erweiterung von VoiceXML um die Steuerung von Videostreams. Im Gegensatz zu WAP werden dabei nicht bloß statische Internetinhalte angezeigt, vielmehr werden die multimedialen Möglichkeiten des Internet in die Telefoniewelt übertragen. Aus Sprachportalen entstanden, sind die 3G-Telefonieanwendungen zudem stärker dialogorientiert als WAP und deshalb deutlich interaktiver, weil die Ein- und Ausgabe bei Sprachdialogen im Unterschied zu Internetanwendungen in Echtzeit ablaufen muß. Statt lediglich Webseiten so darzustellen, daß sie sich für das Handydisplay eignen, wird verstärkt auf die interaktive Gestaltung einer für die Besonderheiten des Mobiltelefons geeigneten Mensch-Maschine-Kommunikation orientiert.

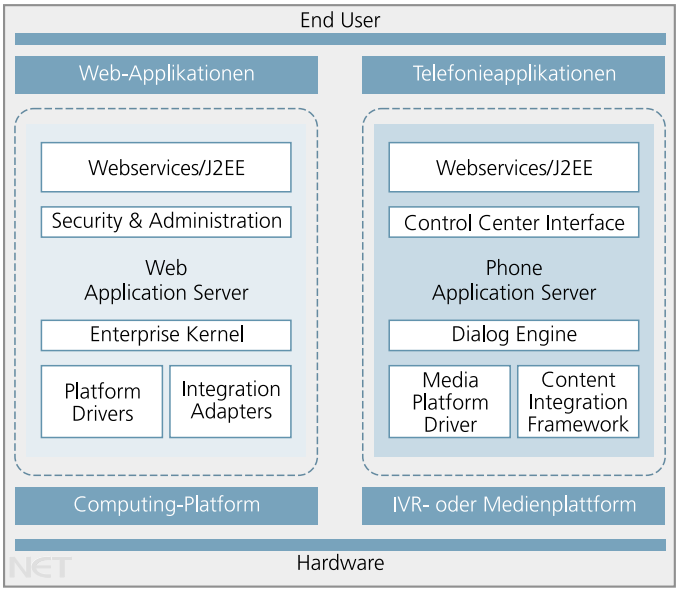


Bild 1: Phone Application Server sind in ihrer Funktion mit Web Application Servern vergleichbar

Das Thema in Kürze

Da sich mittlerweile mit der Markup-Sprache VoiceXML auch Videostreams steuern lassen, können die über Sprachportale gelieferten Sprachinformationen nun auch um multimediale Elemente angereichert und die Benutzer visuell unterstützt werden. Damit werden besonders benutzerfreundliche, dialogorientierte Schnittstellen für Telefonieanwendungen im Mobilfunk der dritten Generation (3G) möglich.

3G-Telefonieportale kombinieren Sprachdialoge mit Video. Dabei bezieht sich der Begriff Video in diesem Zusammenhang nicht allein auf Bewegtbilder, sondern meint die Übermittlung von multimedialen Inhalten, also Bilder, Grafiken, Texteinblendungen, Animationen und Filme, aber auch Geräusche, Musik und gesprochene Informationen per Videostreaming. Alle Inhalte werden dynamisch erzeugt, zu einer Multimediaanwendung kombiniert und über die mobile Datenverbindung in Echtzeit auf das 3G-Endgerät gestreamt. Mit der Ausgabe visueller Inhalte auf das Display kommt zum Modus Sprache ein zusätzlicher Ausgabemodus hinzu. Für die Steuerung und Eingabe kann der User wahlweise die Stimme oder die Handytastatur nutzen. Die Möglichkeit, beliebig zwischen den Ein- und Ausgabemodi zu wechseln, wird als Multimodalität bezeichnet.

Gute Benutzerführung

Phone Application Server (*Bild 1*) dienen bei der Anreicherung der über die Sprachportale gelieferten Sprachinfor-

mationen um multimediale Inhalte als Middleware und stellen die multimediale Benutzeroberfläche bereit. Als Vermittlungsinstanz führen sie Daten aus unterschiedlichen Quellen zusammen und verarbeiten sie so weiter, daß sie sich auf einem UMTS-Gerät abspielen und per Sprach- oder Tasteneingabe steuern lassen. Diese Daten können aus Datenbanken, Business-Intelligence- oder CRM-Systemen, aber auch aus beliebigen Spezialanwendungen wie Banking-, Buchungs- oder Bestellprogrammen kommen.

Welche neuen Möglichkeiten der Benutzerführung Multimedia und Multimodalität bieten, läßt sich am einfachsten an einem Beispiel erklären: Bei einem einfachen Sprachportal muß sich der Zuhörer viele Dinge merken. Das führt zu den bekannten Dialogen wie „Drücken Sie 1 oder sagen Sie 1 für Neukunden, 2 für bestehende Verträge, 3 für Abrechnung" usw. Bis der Sprachcomputer fertig ist mit der Aufzählung, haben viele Kunden die ersten Punkte möglicherweise schon wieder vergessen. Bei Multimediaanwendungen, die mit Phone Application Servern auf Basis von VoiceXML entwickelt wurden, wird der Benutzer davon entlastet. Er muß sich den Dialog nicht mehr merken, sondern könnte die Optionen hören und parallel auf dem Display angezeigt bekommen. Er gewinnt einen Überblick über das Serviceangebot durch die visuelle Unterstützung der Navigation. Je nach Situation oder Vorliebe kann er Steuerungsbefehle sprechen oder Menüpunkte per Tastatur anwählen, Inhalte können dabei audiovisuell gemischt bzw. komplett als reine Audio- oder Displayanwendungen ausgegeben werden.

Lange Zeit hatten Portale, die mit automatischer Spracherzeugung und -erkennung arbeiten, durch die Kinderkrankheiten der Technik mit Akzeptanzproblemen zu kämpfen. Mittlerweile gibt es hier, zumindest bei gut designten Systemen, jedoch deutliche Fortschritte. Unterhaltsam gestaltete Sprachservices bringen genügend Spaßfaktor, um auch als Marketinginstrument eingesetzt werden zu können. So hat der 1. FC Köln im vergangenen Jahr ein sprachgesteuertes Gewinnspiel durchgeführt, bei dem einer der Preise ein Treffen mit dem damaligen Kölner Star Lukas Podolski war. Der Publikumsliebling stellte Fragen zum 1. FC Köln. Durch den Einsatz von Aufzeichnungen der markanten Stimme des Jungkickers wurde die Zielgruppe wirksam angesprochen. Mit Videostreaming könnte daraus nun eine echte UMTS-Anwendung werden. Per VoiceXML ließen sich die Sprachaufzeichnungen mit lippensynchronen Bewegtbildern des Stürmers koppeln. Animierte Kunstfiguren, sog. Avatare, sprechen vor allem die mit Videospielen und multimedialen Internetanwendungen groß gewordene junge Generation der Handynutzer an.



Bild 2: Videotelefonat mit Avatar: Die virtuelle Agentin sagt das Serviceangebot an, gleichzeitig erhält der UMTS-User einen visuellen Überblick über das Menü



Bild 3: Konto-, Telefon- oder Kundennummern lassen sich wahlweise per Sprache oder Tastatur eingeben. Dabei sieht der Anrufer live, was er eingibt, mündliche Rückbestätigungen entfallen (Bilderquelle: VoiceObjects)

Ziel sollten dabei nicht Multimedialität um jeden Preis, sondern je nach Anwendungsszenario der höhere Unterhaltungswert, die bessere Benutzbarkeit im Kundenkontakt oder unternehmensinterner Anwendungen sein. Es geht nicht um Sprachein- oder -ausgabe immer und überall, sondern um die Option und Auswahl.

Das gilt natürlich auch für den Einsatz der Kunstfiguren. Sie müssen sich dem Zweck unterordnen, die Bedienbarkeit zu erleichtern oder angenehmer zu gestalten (*Bilder 2 und 3*). Eine Voraussetzung dafür ist, daß sie gut verständlich und lippensynchron „sprechen". Das wird beim Videostreaming per VoiceXML durch die extrem geringen Latenzzeiten von unter 200 ms erreicht. Diese Werte lassen sich bereits heute schon vor Einführung des Übertragungsstandards HSDPA erzielen.

Kundenbetreuung im UMTS-Zeitalter

Hierbei geht es nicht darum, bestehende und akzeptierte Systeme wie z.B. Fahrplanauskünfte im Internet durch rein sprachbasierte Systeme zu ersetzen, sondern vielmehr soll den Nutzern eine benutzerfreundliche Alternative für den mobilen Einsatz geboten werden. 3G-Telefonportale würde ein Bahnfahrer z.B. dann zu Rate ziehen, wenn der Geschäftstermin länger dauert als erwartet und er später fahren muß als geplant oder wenn er durch Zugverspätungen einen Anschluß verpaßt. Ruft er diese Auskunftssysteme heute mit seinem Handy an, dann zählen sie lediglich die nächsten Verbindungen auf. Zusatzinformationen wie Zuschläge, Zugrestaurant oder andere Details werden nicht unmittelbar angesagt, da das den Dialog überfrachten würde.

Anders bei videogestützten Anwendungen: Der Kunde kann sich auf einen Blick über die Verbindungen informieren und erhält gleichzeitig mehr Detailinformationen. Er kann per Stimme oder Tastatur den gewünschten Zug anwählen und z.B. gleich eine Sitzplatzreservierung vornehmen. Eine Schwäche haben solche gestreamten Informationen allerdings: Sie sind Echtzeitanwendungen und lassen sich nicht speichern. Jedoch ist es möglich, sich eine im Dialog ausgewählte Verbindungsinformation oder ein mobiles Ticket als MMS aufs Handy schicken zu lassen.

Telefonbanking der dritten Generation

Eine Integration unterschiedlicher Informationen und Medien läßt sich auch bei typischen Bankanwendun-

gen realisieren. Ein Beispiel ist der Portfolioüberblick im Daytrading, dem tagesaktuellen An- und Verkauf von Aktien. Hier können Inhalteanbieter die Anzeige des existierenden Portfolios durch aktuelle, dynamisch generierte Grafiken wie Wertpapiercharts oder eine Darstellung der Wertentwicklung des eigenen Portfolios ergänzen. Gegenüber rein sprachbasierten Telefonportalen oder auch Call-Center-Agents haben mobile Multimediaanwendungen einen viel höheren Informationswert. Das UMTS-Handy liefert dem Nutzer so wertvolle Entscheidungshilfen für den Wertpapierhandel und zugleich eine sichere und komfortable Möglichkeit, um z.B. Verkaufsorders abzugeben. Das sind zwar Services, die auch im Internet zur Verfügung stehen. Der Mehrwert besteht in diesem Fall jedoch in der Bequemlichkeit und mobilen Verfügbarkeit.

Mobile Business

Einsatzfelder finden sich zudem im B2B-Umfeld. Das kann z.B. die Abfrage von aktuellen Kundendaten aus CRM-Systemen oder von Informationen zur Produktverfügbarkeit und Preisen aus anderen Backend-Systemen wie z.B. ERP-Anwendungen sein. Hersteller wie SAP haben ein großes Interesse daran, mobile Endnutzer auf ihre Systeme und Anwendungen zugreifen zu lassen und etwa die mobile Suche in Geschäftsdatenbeständen zu erlauben. Da die Mitarbeiter auch unterwegs arbeiten können, steigern die Mobile-Business-Lösungen die Produktivität erheblich. Bislang lassen sich diese Lösungen über Softwareclients auf dem Handy oder browserbasiert realisieren.

VoiceXML-basierte Multimediaanwendungen machen den Zugriff auf Informationen dagegen per Terminal-Emulation verfügbar. Das hat im wesentlichen zwei Vorteile: Die Installation von Clients erfordert einen hohen Aufwand für die IT-Abteilungen. Sie müssen für jede Anwendung je nach mobilem Betriebssystem unterschiedliche Varianten der Clientsoftware auf die Endgeräte aufspielen und verwalten. Jedes Softwareupdate löst aufwendige Prozesse aus. Bei einer VoiceXML-Lösung ist der Zugriff über jedes UMTS-fähige Endgerät möglich, der Wartungsaufwand reduziert sich deutlich. Gegenüber browserbasierten Varianten ist der wichtigste Vorteil die größere Zugriffsgeschwindigkeit.

Zudem setzt die Lösung die Eintrittsbarriere für solche Anwendungen herunter, da sie den unkomplizierten Zugriff über alle möglichen Endgeräte erlaubt. Der mobile Mitarbeiter kann über das Gerät zugreifen, das ihm zur Verfügung steht, und sich für die Ausgabefunktion entscheiden, die gerade paßt. Das kann die Anzeige als Text auf dem Display sein oder die Sprachausgabe beim Autofahren.

Weitere Szenarien im Unternehmensumfeld haben die Entlastung von Personalabteilungen bei Standardverfahren wie Krankmeldungen als Ziel. Dort könnten VoiceXML-basierte Dienste in den Workflow eingebunden werden. Das bedeutet z.B., daß der Mitarbeiter sich nicht nur krankmeldet, sondern gleichzeitig im System Prozesse angestoßen werden, um seine Vertretung zu organisieren.

VoiceXML-Entwicklung

Die Markup-Sprache VoiceXML ist ein Standard des World Wide Web Consortium (W3C). Sie wurde entwickelt, um anwendungsübergreifend austauschbare Bausteine für Sprachdialoge in Telefonieanwendungen zu ermöglichen. Die Integration von Videodiensten war ursprünglich nicht vorgesehen, ist aber ohne Veränderungen möglich. Die aktuell diskutierte Weiterentwicklung des Standards bezieht Multimodalität, d.h. unterschiedliche Ein- und Ausgabemodi, Umgebungen und Endgeräte sowie Rich Media Content ein. Derzeit verfügbare Media Server unterstützen solche Lösungen bereits. Media Server sind Softwarelösungen, die Interactive-Voice-Response-Systeme aus der „alten Telefoniewelt" um multimediale Lösungen erweitern. Sie arbeiten aber zur Zeit bei der Integration von Videoservices mit proprietären Formaten. Ende nächsten Jahres ist mit einem ersten „Video-Draft" des W3C zu rechnen, implementiert werden Lösungen aber bereits heute.

UMTS-Handy als Absatzkanal

Umgesetzt wurden bereits Entertainment-Lösungen. So hat der französische 3G-Netzbetreiber SFR ein Musikportal auf VoiceXML-Basis realisiert, aus dem die Kunden Musik aus unterschiedlichen Jahrzehnten auswählen können. Die Auswahl der Titel erfolgt über die Tasten des Telefons. Der Clou dabei: Die Belegung der Tasten kann sich bei jedem Menüschritt ändern. Sie wird aber jeweils auf dem Display des Handys angezeigt. Der Benutzer muß sich also nicht merken, was ihm gesagt wird, sondern kann es sich anschauen und dann entscheiden.

Völlig neue Möglichkeiten ergeben sich auch über die bereits realisierte Anwendung des SFR hinaus für den Mobile Commerce. Versandhandelsangebote oder Auktionsplattformen wie Ebay haben es sehr viel einfacher, Produkte zu präsentieren, selbst wenn der Benutzer mobil zugreift. Für die derzeit heranwachsende Generation, die schon heute Klingeltöne oder andere handybezogene Produkte und Services bestellt, wird das mobile Einkaufen jeglicher Waren und Dienstleistungen selbstverständlich werden – vor allem, wenn es komfortabel und nicht mit langen Wartezeiten oder Kosten verbunden ist.

Sogar völlig neuartige Formen des Mobile Commerce werden möglich, etwa interaktive Handy-Shopping-Kanäle, eine Kombination aus Teleshopping und mobilem E-Commerce. Zudem wird es neue, crossmediale Strategien des Mobile Marketing geben, die unterschiedliche Elemente und Kanäle verbinden. Denkbar ist beispielsweise, daß die Modemarke X auf Plakaten eine Sonderaktion und eine dazugehörige Telefonnummer bewirbt. Bei Anruf der Nummer kann man über ein kleines Gewinnspiel als Preis z.B. eine mobile Rabattmarke erhalten, die sich dann im Laden einlösen läßt. Den Standort des nächsten Geschäfts erfährt der Anrufer über multimedial gesteuerte Location Based Services. *(bk)*