



Multimodale Anwendungen – Sprachautomation auf allen Kommunikationskanälen

 http://www.crmmanager.de/magazin/artikel_1593_multimodal_sprachautomation.html

Telefon, Internet und Email sind nur einige unserer aktuellen Hilfsmittel in der täglichen Kommunikation mit Kunden, Interessenten und Lieferanten. Das CRM System ist unsere stets aktuelle Datenquelle, wird aber hauptsächlich zur Verwaltung und Dokumentation der Kontakte und Kampagnen genutzt. Wir arbeiten heute ständig auf einer Vielzahl von Kommunikationskanälen und reden immer weniger im direkten persönlichen Gespräch miteinander. Das Telefon als das Hauptmedium der heutigen Kommunikation, hat durch die Email und deren Abwandlungen wie SMS und Instand Messaging, Zuwachs bekommen. Video und weitere Medien stehen in den Startlöchern. Gleiches gilt für die Automation im Kundendialog. Denn Selfservice bedeutet heute viel mehr als nur eine Nachricht auf einem Anrufbeantworter zu hinterlassen.

In vielen Fällen macht es daher im **Bereich Selfservice** Sinn, **mehrere Kanäle** miteinander zu **kombinieren**, um somit die gewünschten Informationen besser und effektiver transportieren zu können. Es gilt die alte Weisheit, ein Bild sagt mehr als tausend Worte und Bilder können von vielen unserer digitalen Begleiter sehr gut dargestellt werden.

Multimodal – Was ist das?

Werden verschiedene Kommunikationskanäle gemischt, so spricht man von Multimodalen Anwendungen. Die einfachste Art einer solchen Anwendung ist eine **Voice-Mail**. Eine Voice-Mail wird per Sprache auf einem Anrufbeantworter hinterlassen und wird per Email an einen Empfänger versendet. Hieraus ergeben sich schon einmal zwei unterschiedliche Kommunikationswege, nämlich der Sprachkanal "Telefon" und der Nachrichtenkanal "Email". Ein dritter Kanal kommt hinzu, wenn die Information zeitgleich per SMS an den Empfänger bzw. sein Mobiltelefon versendet wird.

Dieses simple Beispiel zeigt, wie wir heute bereits Multimodale Anwendungen nutzen. Durch den Mix weiterer Kanäle lassen sich für unterschiedliche Kundengruppen Informationssysteme erstellen, die, durch die individuelle Wahl des Kommunikationskanals, einen sehr hohen Nutzwert haben, da der Kunde selber entscheiden kann, wann und über welchen Weg er Informationen beziehen oder abgeben möchte.

Spracherkennung nutzen

Die Spracherkennung findet seit einigen Jahren immer mehr Beachtung und ermöglicht wunderbare Lösungen, die einem sehr hohen Faktor an Automation erlauben. Der Computer versteht endlich was der Benutzer sagt und kann so Datenbanken in Echtzeit abfragen, um dem Benutzer schnell und effektiv zu dienen. Die Kommunikation a la "Captian Kirk von der Enterprise", der einfach und natürlich sprachlich mit der Maschine redet ist keine Illusion mehr, sondern findet in einer Vielzahl von Fällen Ihre Anwendung.

Seit Jahren gibt es Systeme, die den Kontostand am Telefon ansagen, die Termine der Filme im örtlichen Kino per Telefon vorlesen oder auch eine Bahnverbindung

am Telefon wiedergeben. Viel interessanter wird es aber, wenn nun die Kommunikationskanäle gemischt werden, denn so lässt sich auch unterschiedlicher Inhalt transportieren. Am Telefon kann nur gesprochen werden, aber wie bereits erwähnt, sagt ein Bild oder ein Video viel mehr als tausend Worte. Die heutigen Telefone, gerade die aktuellen Mobiltelefone, erlauben mittlerweile die Nutzung von Video.

So lassen sich **Lösungen erdenken**, die bei einem Anruf bei der Kinohotline einen Ausschnitt aus dem neuen Kinofilm "Harry Potter" sofort auf das Handy bringen. Das ist Kundennutzen! Im Nachspann wird automatisch gefragt wie viele Karten auf die Handynummer zur Abholung hinterlegt werden sollen. Der Kunde kann ganz einfach per Sprachdialog agieren. Ein einfacher Satz wie "Reservieren Sie bitte 4 Karte für Harry Potter um 21.00 Uhr" stößt den Bestellprozess an, der natürlich die Bestellung sofort per SMS bestätigt. Diese Lösung ist so noch nicht im Einsatz, aber eine Vielzahl weiterer nützlicher Multimodaler Anwendung finden sich heute bereits in Deutschland.

S- und U-Bahn per Sprache abfragen

Der **MVV in München** hat ein System zur komfortablen **Abfrage aller S- und U-Bahnen wie auch Busverbindungen** für den Großraum München erstellt, welches auf einfache Multimodale Funktionen setzt. Das System zeichnet sich durch eine klare und einfache Benutzung aus und eignet sich daher sehr gut als Beispiel. Ein Anruf auf der Festnetzrufnummer (Testrufnummer: 089 / 41 42 43 44) ermöglicht dem Benutzer die Angabe seiner Route durch Nennung der Abfahrt- und Zielhaltestelle, wie auch der Uhrzeit. Der geführte Sprachdialog ermöglicht eine gute Dateneingabe und hält Optionen wie die gewünschte Ankunftszeit bereit. Der Clou ist, dass Sie nach Erhalt der Information am Telefon diese zudem als SMS oder als Fax bekommen können. So haben Sie jederzeit die Möglichkeit diese nochmals anzuschauen um zu erkennen, an welchen Haltestellen Sie umsteigen müssen oder ob beim Umsteigen noch Zeit für einen Imbiss bleibt. Die Multimodale Nutzung bezieht sich hier auf die einfache Eingabe per Sprache und die Ausgabe in ganz anderen Kanälen.

Entertainment fast ohne Tasten

Es lassen sich auch ganz andere Applikationen erstellen, die wiederum andere Kommunikationswege nutzen. Das eine Sprachanwendung nicht langweilig sondern eher sehr interaktiv sein kann, dabei sogar eine Verbindung zum Internet bzw. unserem beliebten Browser herstellt, eröffnet völlig neue Horizonte. So hat das Münchner Unternehmen **Telenet** zu seinem 25-jährigen Firmenjubiläum die Applikation "**Oma Anneliese**" erstellt. In dieser multimodalen Anwendung verschmelzen das Internet in Form des Kommunikationskanals "Browser" und der Kanal "Telefon mit Spracherkennung" zu einer ansehnlichen Lösung.

**Oma Anneliese –
das sprechende Kreuzworträtsel**

Lösungsfragen

- 1 Vorderer Teil eines Schiffes
- 2 Südlicher Kontinent
- 3 Sehorgan
- 4 Trinkgefäß
- 5 Huftier
- 6 Zitrusfrucht

Lösung T R A B

Es gilt gemeinsam mit der fiktiven "Oma Anneliese" ein Kreuzworträtsel auszufüllen ohne dabei die Tastatur oder einen Stift zu nutzen. Nur per Stimme füllen sich die Zeilen und Spalten am Bildschirm sichtbar vor einem. Ein Anruf auf der Ortsrufnummer der "Oma Anneliese" stellt per Eingabe einer Nummer die Verbindung zwischen dem Browser und dem Anruf her. Danach geht alles wie von Geisterhand geführt. Der frische Dialog mit der Oma wurde von den VUI-Designern der Telenet erstellt und von den Sounddesignern der Comevis aus Köln vertont.

The image shows a screenshot of a website interface. On the left, there is a blue header with the text "Herzlichen Glückwunsch, Sie haben das Rätsel gelöst!". Below this is a registration form with fields for "Nachname*", "Vorname*", "Firmenname", "Straße, Haus-Nr.", "PLZ, Stadt", "Land", "E-Mail-Adresse*", and "Telefon". There are two checkboxes for opting into newsletters. At the bottom of the form is a blue "Abschicken" button. On the right, there is a crossword puzzle grid. The words "AUSTRALIEN" and "GEGESSEL" are filled in horizontally, and "RANG" is filled in vertically. Numbers 3, 4, 5, and 6 are placed in the starting squares of the words. At the bottom of the grid, the word "STRASSBURG" is written in a large, blue, outlined font.

Die freundliche Oma hat immer wieder motivierende Worte zur Verfügung, damit der Benutzer die richtigen Begriffe im Kreuzworträtsel findet. Findet der Anrufer dann das Lösungswort, so winkt auch noch jeden Monat ein Preisverleih. Die Mischung aus Spracherkennung, frischer Benutzerführung und der Visualisierung der erkannten Worte auf dem Bildschirm zeigen in welche Richtung sich die Lösungen entwickeln werden.

Neue Anwendungen – aber richtig

Die Beispiele geben erste Ideen für den Einsatz von Multimodalen Anwendungen in der Praxis. In diesem Markt ist noch viel Potential versteckt, denn die immer weitere Verbreitung der unterschiedlichsten Endgeräte und die Nutzung von Sprachtechnologie erlaubt permanent weitere Möglichkeiten.

Nur ist **jeder Anwendungsfall bekanntlich unterschiedlich**, basiert auf völlig verschiedenen Datenbanken und verfolgt unterschiedliche Ziele. Je nachdem ob eine Lösung im Bereich der Finanzwirtschaft, dem Entertainment oder im Bereich einer Prozesskette in einem Produzierenden Unternehmen entstehen soll, muss ein Ansatz gefunden werden, der dem potentiellen Nutzer wirklich hilft.

Dies fängt bei der passenden Wortwahl an und geht über die richtige grafische Schnittstelle bis hin zur Auswahl der passenden Endgeräte, die der Nutzer zur Verfügung hat. Allein aus dieser **Vielzahl von Voraussetzungen** sei darauf hingewiesen, dass Sie bei einem solchen Projekt auf Wissen von Experten zurückgreifen sollten. Mittlerweile finden sich in Deutschland verschiedene **Voice User Interface Designer** (kurz VUI-Designer), die mit einem untechnischen Ansatz an den Start gehen und Unternehmen bei der Erstellung einer Anwendung im linguistischen Bereich helfen.

Wie vor einigen Jahren die Webdesigner neu auf der Bildfläche erschienen sind, so ist es im Fall einer Sprachanwendung der VUI-Designer, der diese komplexe und

multimodale Welt versteht und bei allen Überlegungen versuchen soll den Kunden oder Nutzer der Anwendung im Fokus zu halten. Er hat die Aufgabe, zu verhindern, dass der neue Anwendungsfall ein zu technisches Kundenverhinderungssystem wird. Nur wenn der Benutzer Ihre neue Anwendung gern verwendet, wird er es wieder tun und so nachhaltig für eine Auslastung und den Erfolg der Anwendung sorgen.

Standards

Das Multimodale Anwendungen keine technische Spielerei sind, zeigt auch das Interesse des **World Wide Web Consortiums** (W3C), welches Standards zum Austausch von Informationen Multimodaler Anwendungen definiert. **EMMA** heißt eine Beschreibungssprache die auf XML basiert. Diese "Extensible MultiModal Annotation markup language" dient als Grundlage, um alle nur erdenklichen Inhalte multimodal miteinander zu verknüpfen. Die Definition dieses Standards befindet sich in dem Status des "Working Draft" und wird sicher bald finalisiert werden. Weitere Informationen finden Sie [hier](#).

Fazit

Der richtige Mix in der Selbstbedienung macht den feinen Unterschied. Mit Multimodalen Lösungen erreichen Sie den Kunden auf allen erdenklichen Kanälen und können Ihre Information somit optimal transportieren. Im Vorfeld gilt es allerdings genau die notwendigen Schnittstellen zu untersuchen. Hierbei sollten Sie auf die technischen Schnittstellen wie Datenstrukturen und verwendete Technologie genauso viel Zeit investieren, wie in die didaktischen und linguistischen Schnittstellen, die die Interaktion mit dem Kunden beschreiben.

voice compass 2007



Sie planen eine Sprachanwendung oder eine Multimodale Anwendung und suchen die passenden Dienstleister oder Produkthanbieter wie auch Empfehlungen aus der Praxis. Detlev Artelt, der Autor dieses Artikels, hat das Standardwerk zu diesem Fachbereich erstellt und beschreibt in dem Buch "[voice compass 2007](#)" in aller Ausführlichkeit die Grundlagen für Sprachanwendungen. Zusammen mit 4 Co-Autoren wird der Anbietermarkt geordnet und das Thema Sprachanwendung entmystifiziert.

[Weitere Informationen zum Buch...](#)

Erschienen: 09/2007

Autor: Detlev Artelt



Detlev Artelt ist Geschäftsführer der aixvox GmbH in Aachen und namhafter Autor verschiedener Fachbeiträge und Marktstudien auf dem Gebiet datenbankgestützter Computersprachsysteme in Verbindung mit Spracherkennung und -synthese.