Automatisierte Vermittlungen mit Spracherkennung: Eine Akzeptanzanalyse

Diplomarbeit

zur Erlangung des Grades einer Diplom-Ökonomin der

Wirtschaftswissenschaftlichen Fakultät der Leibniz Universität Hannover

vorgelegt von: cand. oec. Maren Matusch

[redacted]

Erstprüfer: Prof. Dr. Michael H. Breitner

Hannover, den 04.03.2008

Inhaltsverzeichnis

Abkürzungsverzeichnis....3

Abbildungsverzeichnis....5

1 Einleitung....6

2 Einführung in das Thema Spracherkennung....10

2.1 Wichtige Begriffe....10
2.2 Anwendungsbeispiele mit Spracherkennung....14

3 Die Automatische Vermittlung....17

3.1 Einordnung in den Kontext....17
3.2 Bestandteile einer sprachbasierten Anwendung....18
3.3 Erklärung des Ablaufs einer Vermittlung....20
3.4 Einsatzmöglichkeiten für eine Automatische Vermittlung....22
3.4.1 Peterzahlt.de....22
3.4.2 Sprachbasierte Bewerberhotline für Call Center Mitarbeiter....23
3.4.3 Self Service Portal der Sparda Bank West....24

4 Eigenentwicklung oder Hosting – Eine Überlegung wert....25

4.1 Allgemeine Beschreibung....25
4.2 Ein Rechenbeispiel....25
4.3 Vor- und Nachteile einer Hosting-Lösung....28

5 Anforderungen an das System....29

6 Klare Vorteile von automatischen Vermittlungsdiensten....32

6.1 Die Sicht der Betreiber....32
6.2 Aus Kundensicht....35

7 Kritische Erfolgsfaktoren....37

8 Das Thema: „Akzeptanz von Automatischen Vermittlungen“....39

8.1 Kanäle und geeignete Instrumente für die Befragung....39
8.2 Art der Fragestellung....40
8.3 Teilnehmer und Veröffentlichung....41
8.4 Das Instrument LimeSurvey für die Befragung im Internet....42
8.5 Fragestellungen und Details....43
8.6 Auswertung der Internetbefragung....50
8.6.1 Allgemeine Informationen....50
8.6.2 Bekanntheit des Begriffs....51
8.6.3 Benutzung einer Automatischen Vermittlung....51
8.6.4 Einsatz einer Automatischen Vermittlung....55
8.7 Telefonische Befragung per Sprachcomputer....61
.. 8.8 Publikumsbefragung beim Kongress Call Center Science....62

9 Implikationen....63

10 Zukunftsaussichten / Ausblick....66

11 Literaturverzeichnis....71

1 Einleitung

Der Computer als allgegenwärtiger Gebrauchsgegenstand ist heute weder aus dem privaten Bereich, noch aus dem Unternehmenssektor, wegzudenken. Dieses wird durch die deutliche Tendenz bestätigt, dass der Anteil jener Personen, die einen Computer nutzen, stetig ansteigt. Einer Studie des Bundesverbands Informationswirtschaft, Telekommunikation und neue Medien e. V. zufolge, betrug der Anteil an Personen in Deutschland, die im privaten Bereich einen PC nutzen, im Jahr 2007 70 Prozent. 2003 waren es lediglich 56 Prozent. Seitdem hat sich der Anteil privater PC-Nutzer stetig erhöht.[1] Dieser Trend zeichnet sich auch bei Angestellten ab. Während 2003 nur 44 Prozent der in Unternehmen beschäftigten Personen angaben, am Arbeitsplatz einen Computer zu nutzen, sind es im Jahr 2007 schon 61 Prozent. Und obwohl dieser Wert niedriger ist als bei den privaten Nutzern, ist auch hier ein kontinuierlicher Anstieg der Nutzerzahlen zu beobachten.[2] Es wird ersichtlich, dass sich immer mehr Menschen, sei es im beruflichen oder privaten Alltag, mit der Nutzung eines PC auseinander setzen.

Eine weitere bedeutsame Entwicklung stellt die Automation vieler Prozesse dar. Besonders im Unternehmenssektor verändern sich viele Abläufe, da sie teilweise oder vollständig automatisiert werden. Für das jeweilige Unternehmen resultieren aus diesen Maßnahmen der Automatisierung häufig Einsparungen in Form von Personal und Kosten.

Aus diesen Gründen wird der Einsatz von Spracherkennungssystemen für Unternehmen zunehmend interessant und die Verbreitung solcher Systeme nimmt immer mehr zu. Die Gruppe der Endkunden stellt eine wichtige Zielgruppe dar, deren Wünsche und Bedürfnisse zu befriedigen sind. Aber auch die Kommunikation zwischen Unternehmen wird immer häufiger durch den Einsatz von Spracherkennung geprägt, so dass Mitarbeiter anderer Firmen als Kunden fungieren. Ein sehr entscheidender Faktor, der den Erfolg einer Organisation maßgeblich beeinflusst, liegt in der Kundenakzeptanz der sprachbasierten Anwendungen.

---

1 Pols, A. [2008a]
2 Pols, A. [2008b]

Bild 1: "Automatisierte Prozesse – Last oder Lust?", aixvox GmbH

Wie ausgeprägt ist die Akzeptanz der Benutzer bereits? Gibt es Möglichkeiten, diese noch zu erhöhen, und gleichzeitig vorhandene Schwachstellen in den Anwendungen zu beheben?

Über viele Jahre hinweg war der Begriff Spracherkennung „nicht greifbar“. Obwohl bereits in den sechziger Jahren die Forschungen im Bereich Spracherkennung begannen, wurde erstmalig auf der CeBIT 1991 das System TANGORA 4 als deutsche Version von IBM vorgestellt. Der Wortschatz umfasste zum damaligen Zeitpunkt 20.000 bis 30.000 Begriffe. Eingesetzt wurde das System im medizinischen und juristischen Bereich sowie für allgemeine Korrespondenz in Unternehmen. Die Demonstration fand in einem abgeschirmten Raum statt, so dass die Spracherkennung nicht durch Hintergrundgeräusche beeinflusst wurde.

Nach dieser Entwicklung gab es immer wieder neue Anwendungen, die auf technologischer Basis den bisherigen voraus waren. Kurze nach der CeBIT im Jahr 1994 wurde die deutsche Variante eines Diktiersystems der Öffentlichkeit vorgestellt. Dieses

System konnte auf der CeBIT 1995 bereits ohne Abschirmung gegen Hintergrundgeräusche präsentiert werden. Andere Unternehmen, die ebenfalls in den achtziger Jahren die Entwicklungen voran trieben, sind Dragon Systems und Philips Dictation Systems.[3]

Heute werden Spracherkennungssysteme in immer mehr Wirtschaftsbereichen sinnvoll eingesetzt und Neuentwicklungen drängen schneller als in der Vergangenheit auf den Markt. Auch Endverbraucher kommen vermehrt mit automatisierten Anwendungen in Kontakt, welche auf Basis von Sprache funktionieren. Verschiedenste Lösungen wurden bereits realisiert (Self Service, Auskunftsdienste, Bestellservice, Sprecherverifikation etc.) und ständig kommen neue Ideen hinzu. Aufgrund der Aktualität des Themas ist deshalb die Fragestellung interessant, wie hoch die Akzeptanz dieser Systeme (speziell automatischer Vermittlungsdienste) bei den Benutzern, insbesondere im B2B-Sektor, ist.

Unternehmen, die Überlegungen zum Einsatz eines Spracherkennungssystems anstellen, müssen sich zunächst mit der Frage auseinander setzen, ob die ausgewählten Funktionen überhaupt automatisierbar sind. Falls dies der Fall ist, kann eine Entscheidung über die geeignete Anwendung getroffen werden. Die Wahl fällt entweder zugunsten eines Mensch-Maschine Mix oder einer rein auf Sprache basierenden Lösung.

Nach einer allgemeinen Einführung in das Thema mit Begriffserklärungen und einer Beschreibung einiger Einsatzbeispiele folgt die Einordung in den Kontext, eine kurze Darstellung der Komponenten eines Voice-Portals und des Ablaufs einer Automatischen Vermittlung. Anschließend werden Einsatzmöglichkeiten von sprachbasierten Vermittlungsdiensten beschrieben und der Unterschied zwischen dem Eigenbetrieb und einer Mietlösung anhand eines Rechenbeispiels aufgezeigt. In Bezug auf die genannten Aspekte folgt die Erläuterung von Anforderungen, Vorteilen und kritischen Faktoren bei Einsatz einer Automatischen Vermittlung. Die anschließende Befragung an einen definierten Kreis von Unternehmen soll

3 O. V., Linguatec [2007]

Aufschluss geben, inwieweit die Akzeptanz bei den Usern im B2B-Bereich bzw. einsetzenden Unternehmen eines solchen Systems vorhanden ist. Zudem wird versucht, aus den Ergebnissen Handlungsempfehlungen abzuleiten, die zu einer höheren Akzeptanz bei den Benutzern führen können. Die Arbeit schließt mit einigen Zukunftsaussichten ab.

10 Zukunftsaussichten / Ausblick

In den letzten 50 Jahren hat sich der Einsatz von Spracherkennung von einer Idee zur Realität entwickelt. Es bieten sich durch den Einsatz von automatisierten Sprachanwendungen heute neue ökonomische und effektive Lösungen für Unternehmen, die Kundenzufriedenheit zu erhöhen und gleichzeitig fixe Kosten zu senken.[71]

Allerdings gibt es noch immer Konsumenten, die gegenüber der Benutzung von Sprachanwendungen voreingenommen sind. Sie vermuten einen höheren Zeitaufwand, bis sie mit dem gewünschten Gesprächspartner verbunden werden, da das System sie nicht beim ersten Nennen des Namens versteht. Hinzu kommt, dass manche Personen sich gänzlich scheuen, eine Automatische Vermittlung anstelle einer persönlichen Weiterleitung auszuprobieren.[72]

Eine Aufgabe der Unternehmen besteht darin, die genannten Vorurteile zu entkräften, indem sie ihren Kunden beweisen, dass ihre Vorbehalte gegenüber der Nutzung einer solchen Anwendung unberechtigt sind.

Immer mehr Unternehmen nehmen deshalb eine Umstellung auf ein automatisiertes System vor. Wichtige Gründe, die für diese Entwicklung sprechen, sind der steigende Kostendruck, wachsende Anruferzahlen, ebenso wie das Bedürfnis, Kunden durch besseren Service zu binden. Dieser Trend wird sich auch in der Zukunft fortsetzen.[73] Im Vergleich zu sprachbasierten Systemen haben ältere, auf Tonwahl basierende Anwendungen diverse Nachteile. Die Weiterleitung zu einem persönlichen Agenten nimmt mehr Zeit in Anspruch, die Menüstruktur ist sehr umfangreich und Anrufer werden mit langen Wartezeiten konfrontiert.[74]

Automatische Sprachdialogsysteme sind der erste Kontakt mit einem Unternehmen und prägen den Eindruck der Kunden stark. Hier werden Anrufe entgegen genommen und an entsprechende Nebenstellen weitergeleitet. Für die Agenten bedeutet dieser Umstand eine

71 Hong, D. [2006, S. 19]
72 Peissner, M., Sell, D., Steimel, B. [2006]
73 Heßdörfer, J. [2004]
74 O. V., Nuance [2005a, S. 6]

enorme Entlastung, da sie sich nicht mehr mit der Tätigkeit der reinen Vermittlung befassen müssen.[75]

Ein Auto Attendant sollte jedoch nicht als Ersatz für den Einsatz von Agenten gesehen werden. Vielmehr stellt eine solche Anwendung eine sinnvolle Ergänzung zu persönlichen Ansprechpartnern im Call Center eines Unternehmens dar. Zusammen wirkt das Erscheinungsbild des Unternehmens nach außen professionell und „rund“.[76]

Bei Betrachtung der Betreiberseite fällt auf, dass heutige Anwendungen im Voice-Business noch immer Limitationen unterliegen. Bestehende, auf Sprache basierende Systeme sind nicht in der Lage, einen unbegrenzten Wortschatz der Kunden zu verstehen. Diese erwarten jedoch von den Maschinen, dass sie wie ein Mensch kommunizieren können und Antworten nicht auf vorher definierten Worten oder Wortfolgen beruhen müssen. Bisher entwickelte Sprachanwendungen erfüllen den gewünschten Zweck, sofern die benutzten Informationsinhalte begrenzt und der Dialogablauf klar strukturiert werden kann.

In den Anwendungen selbst gibt es die Einschränkung, dass es trotz kontinuierlicher Weiterentwicklung kaum möglich ist, Folgen alphanumerischer Zeichen problemlos zu erkennen. Zum Beispiel erweist sich die richtige Identifikation von Email-Adressen oder die Unterscheidung zwischen den Buchstaben „f“ und „s“ schwierig. Als Lösung, welche bereits Einsatz findet, wird eine Datenbasis mit zuvor festgelegten Wörtern verwendet, die während eines Anrufs vom System mit den gesprochenen Zeichenfolgen verglichen wird, um die Erkennung zu realisieren.[77]

Es existiert ein weiteres Problem, das auch in der nahen Zukunft kaum zu beheben sein wird. Eine Maschine wie die Automatische Vermittlung kann bestimmte Ausdrucksweisen, welche neben der gesprochenen Sprache in einem Gespräch auftreten, nicht verstehen. Dazu zählen Betonungen, die bestimmte Inhalte

75 O. V., PC Magazine [2007a]
76 Eidsvik, B. [2007]
77 O. V., Envox Worldwide [2007]

hervorheben sollen, genauso wie zwischenmenschliche Artikulationen in Form von Gesten, Blicken, kurzen Lauten und nicht ausgesprochenen Worten. Anders führen auch laute Hintergrundgeräusche zu Problemen. Eine Anwendung versteht die gesprochenen Worte und Ausdrucksweisen der Gesprächspartner nicht. Somit werden die Bedeutungen von Gesprächsinhalten nicht erkannt. Beispielsweise erkennt ein System nicht, über welche Themen bei einer Feier mit lauter Geräuschkulisse gesprochen wird. Menschen sind dazu in der Lage.[78]

Die bisherigen Entwicklungen im Bereich Spracherkennung lassen noch genügend Raum für Verbesserungen. Ziele sollten sein, die Unterscheidung und Erkennung von Dialekten weiter zu verfeinern und die Technik dahingehend zu entwickeln, dass Hintergrundgeräusche während eines Anrufs mühelos heraus gefiltert werden können. Außerdem ist eine ständige Anpassung der Benutzeroberfläche für die sprachbasierten Anwendungen an sich ändernde Kundenbedürfnisse notwendig. Nur mit Hilfe solcher und weiterer Maßnahmen resultiert daraus eine Steigerung der Bedienerfreundlichkeit.

Mit Sicht auf den zukünftigen Nutzer einer Anwendung werden „neue“ Voice User Interface Designer die Chance bekommen, ihr Können unter Beweis zu stellen. Ihr Fokus liegt nicht primär auf der Technik. Sie untersuchen den sprachwissenschaftlichen Teil einer Anwendung, und ihr Ziel ist es, immer den zukünftigen Kunden im Blick, Anreize zu schaffen, damit dieser das Sprachportal erneut, oder besser sogar dauerhaft, nutzt.[79]

Bereits heute sind einige Arten multimodaler Anwendungen im Einsatz. Weitere folgen in den nächsten Jahren. Als multimodal bezeichnet werden Systeme, die mindestens zwei Kommunikationskanäle miteinander kombinieren. Die Eingabe erfolgt per Sprache, die Ausgabe über andere Kanäle. Es kann somit unter anderem eine Verknüpfung zwischen Telefon, E-Mail, SMS und Video stattfinden. Der entscheidende Vorteil für den Nutzer liegt in der Möglichkeit, eigenständig die Entscheidung treffen zu können, welche Kanäle er zu welcher Zeit zur Informationsbeschaffung verwenden möchte.

Als gutes Beispiel dient an dieser Stelle das System des Münchner Verkehrs- und Tarifverbunds. Ein Sprachportal unter der Rufnummer 089 / 41 42 43 44 nimmt den Anruf

78 O. V., Linguatec [2007]
79 Artelt, D. [2007b]

entgegen. Nach Eingabe der Abfahrts- und Ankunftshaltestelle sowie der gewünschten Uhrzeit per Sprache erhält der Kunde die entsprechenden Informationen am Telefon. Zusätzlich wird ihm die Datenübermittlung per SMS und Fax angeboten. Für den Benutzer besteht so die Möglichkeit, nach Wahl die Informationen zu einem späteren Zeitpunkt erneut zu betrachten, um die angestrebte Fahrt im Detail zu planen.

Ein weiteres Beispiel für eine multimodale Anwendung ist die Voice-Mail. Hier hinterlässt der Anrufer eine Sprachnachricht auf dem Anrufbeantworter des gewünschten Gesprächspartners. Die Nachricht wird per E-Mail an diese Person weitergeleitet. Alternativ oder als Zusatz kann eine SMS mit der Information versendet werden, dass eine neue Nachricht eingegangen ist.[80] Zunächst wird der Sprach-, danach der Nachrichten- und Datenkanal, verwendet.

Es wird deutlich, dass durch die Realisierung multimodaler Prozesse weitere Potenziale für Unternehmen sowie Nutzer entstehen. Die zunehmende Verbreitung verschiedener Endgeräte und neue Einsatzmöglichkeiten der Sprachtechnologie treiben diese Entwicklung in der Zukunft weiter voran.

Als neue Entwicklung gelten auf Sprache basierende Anwendungen, die in der Lage sind, während eines Anrufs die emotionale Verfassung des jeweiligen Kunden erkennen zu können. Auch die Einordnung in Alter und Geschlecht sowie die Bestimmung der Sprache ist möglich. Zunächst werden dafür online Daten erhoben, anhand welcher eine Klassifikation der Kunden stattfindet. Die Realisierung dieser Lösung erlaubt somit eine individuelle Ansprache eines jeden Anrufers und den darauf angepassten Gesprächsverlauf. Für Kunden erhöht sich die mit dem Anruf verbundene Servicequalität.

Entwickelt wurde ein System von Sympalog Voice Solutions GmbH. Auf den „VoiceDays 2007“ in Bonn erhielt das Unternehmen dafür den „Best Innovation Award“. Die Anwendung wird bereits vom Mobilfunkanbieter T-Mobile in der

80 Artelt, D. [2007b]

Kundenhotline eingesetzt.[81]

Die Arbeit zeigt, dass Sprache heute und vermutlich auch in Zukunft das alleinige Kommunikationsmittel ist, welches keinen Grenzen und Einschränkungen unterliegt. Weitere Entwicklungen und der Fortschritt der Technologien werden neue Anwendungen auf den Markt bringen. Diese Anwendungen werden den Nutzen für Kunden nachhaltig erhöhen, und den einsetzenden Unternehmen, neben den genannten, weitere Vorteile eröffnen.

81 O. V., Sympalog [2007b]