

Interaktive Sprachdialogsysteme: Dienst am Kunden oder Terror aus der Box?

Volker Kraft
Mannesmann Mobilfunk GmbH, Düsseldorf
volker.kraft@d2mannesmann.de

Kurzfassung

Nachdem die Grundfunktionalität „Telefonie“ durch die meisten Mobilfunknetze flächendeckend sichergestellt ist, differenzieren sich die Betreiber wesentlich über ihre Mehrwertdienste. Dem Kunden eröffnen sich hierdurch immer neue Möglichkeiten zur Kommunikation, Information und Transaktion. Rasant steigende Teilnehmerzahlen, stetiger Preisverfall und Rationalisierungsdruck zwingen auch bei den sprachbasierten Diensten (*Interactive Voice Response, IVR*) zur weitgehenden Automatisierung. Dieser Beitrag beschreibt den Status und die zu erwartende Entwicklung von sprachbasierten Diensten und Dienste-Plattformen (*Voice Response Units, VRU*), insbesondere im Hinblick auf die Sicherstellung der Dienste-Qualität (*Quality-of-Service*).

I. Einleitung

Eine Reportage über automatisierte Telefondienste wurde einmal mit „Terror aus der Box“ betitelt. Und wer kennt nicht auch aus eigener Erfahrung den Frust beim Anruf einer der vielen, oft schlecht gemachten Kino-Hotlines?

Demgegenüber sind in den USA zahlreiche Telefondienste äußerst erfolgreich, die teilweise mehr als eine Millionen Anrufer pro Tag bedienen. Typische Applikationen mit hohem Nutzungsaufkommen sind Telebanking, Bestell- und Buchungsdienste sowie Informationsdienste.

Ein weiterer Anwendungsbereich ist die Automatisierung von Call Centern. Bei Kosten von über 2 DM pro Operator-Gesprächsminute amortisieren sich die Investitionen meist in kürzester Zeit. Telefonmitarbeiter werden zudem von Routinearbeiten entlastet. Den Kunden kann rund-um-die-Uhr ein preiswerter Dienst ohne frustrierende Warteschlange und mit konstanter Qualität geboten werden. Die Gartner Gruppe erwartet dass ab 2003 von allen automatisierten Telefonleitungen nahezu jede dritte mit Spracherkennung ausgestattet sein wird [1].

Um den heute über 10 Millionen D2-Teilnehmern überall und jederzeit sofortigen Zugriff auf die wichtigsten Dienstleistungen zu bieten, verfügt D2 über einige hundert Telefonarbeitsplätzen in verteilten Call Centern sowie über Sprachdialogsysteme mit insgesamt über tausend Sprachleitungen.

Wie nützlich ist aber nun die Technologie für den Kunden? Wie bedienfreundlich ist der Dialog? Wie hoch ist letztendlich die Akzeptanz? Soviel ist sicher: Neben dem Nutzwert und dem Preis einer Applikation spielt die Bedienschnittstelle eine wichtige Schlüsselrolle für den Markterfolg.

II. Dienste

Die Grundlage des D2-Dienste-Portfolios [2] bilden - wie in jedem Mobilfunknetz - die GSM-Basisdienste wie Sprachtelefonie, der Kurznachrichtendienst sowie Fax- und Datendienste. Die Zusatzdienste bieten Komfortfunktionen wie Rufumleitung, Makeln oder Konferenz.

Die meistgenutzte IVR-Applikation ist heute die D2-Mailbox, der intelligente Anrufbeantworter im Netz. Außerdem von Bedeutung ist der PrePaid-Service D2-CallYa, wobei über eine VRU Telefongutscheine aufgebucht und Kontoinformationen abgerufen werden. Sowohl die Sprach-Mailbox als auch die CallYa-Dienste werden heute noch ausschließlich mit der Telefontastatur über DTMF-Töne gesteuert. Die Systemausgaben erfolgen über die Wiedergabe von gespeicherten Sprachansagen.

Im Bereich der D2-Kundenbetreuung werden bereits viele Standardfragen in einem automatisierten Einstiegsdialog beantwortet, ohne dass der Kunde an einen Telefonmitarbeiter

durchgestellt wird. Auch dieser Dialog sowie ein automatisierter Faxabrufdienst sind – noch – DTMF-basiert.

Informationsdienste wie beispielsweise die D2-Auskunft, D2-Hotelreservierung oder D2-Wetterinfo werden zunehmend vom Operatorbetrieb auf IVR umgestellt. Weitere Dienste, die bereits über VRUs abgewickelt werden, sind D2-QuickCheck (Kontostandsabfrage), D2-BestCity (Buchung und Einstellung für Ortsgespräche) sowie die D2-ServiceWorld mit Freundschaftwerbung, Gewinnspielen etc.

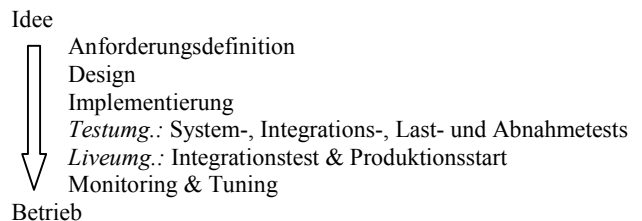
Die Nutzung heutiger Dienste ist bereits über unterschiedliche Kanäle möglich, die je nach Anwendungssituation und Kundenpräferenz gewählt werden:

- natürliche Sprache → Operator
- IVR / DTMF → Voice Response Unit
- IVR / Sprachkommandos → Voice Response Unit
- Daten / Kurznachrichten → Dienstesystem
- Internet / WAP → Dienstesystem
- Internet / Web-Services → Dienstesystem

III. Entwicklungsprozess

Die Entwicklung eines neuen Dienstes dient in der Regel zur strategischen Erweiterung des Portfolios oder zur Rationalisierung interner Prozesse. Auftraggeber ist meist das (interne) Marketing, Endkunden sind in der Ideenphase meist noch nicht beteiligt. Somit müssen die Nutzer bezüglich ihrer Anforderungen und ihres Nutzungsprofils modelliert werden.

Der Realisierungsprozess eines sprachbasierten Dienstes folgt dabei der klassischen Software-Entwicklung mit den Phasen



Besonders hoher Aufwand wird dabei auf die funktionalen Tests verwendet, die schwerpunktmäßig in einem separaten Testnetz stattfinden. Schließlich müssen die Tests den hohen Ansprüchen der Qualitätssicherung genügen und der Abnahme des Produktes durch den Auftraggeber dienen.

Die Bedienschnittstelle einschließlich aller Fehlerfälle wird sowohl vom Hersteller als auch vom Auftraggeber vollständig auf korrekte Funktionalität getestet. Ist der Dienst einmal im Markt eingeführt, sind nachträgliche Korrekturen an der Bedienschnittstelle nur schwer möglich. Außerdem ist mit jeder Änderung ein besonders hoher Aufwand verbunden, sobald auch Abläufe im Netzbetrieb oder in der Kundenbetreuung betroffen sind.

Unter diesen Aspekten ist ein später Kundenkontakt, der erstmalig nach der Produkteinführung ins Feld zustande kommt, besonders bei neuartigen Anwendungen oder komplexen Dialogen als Risiko zu werten. Die Vorgehensweise bei der Entwicklung neuer Dienste sollte daher wie unten diskutiert angepaßt werden.

IV. Technologie

Die Dienste werden in Form von Applikationssoftware implementiert. Meist können mehrere unterschiedliche Applikationen parallel auf einer Dienste-Plattform betrieben werden. Diese sogenannten *Sprachdialogsysteme* bestehen aus vernetzten EDV-Systemen und sind modular aufgebaut. Zu den

wichtigsten Plattform-Eigenschaften zählen Skalierbarkeit, Erweiterbarkeit und Ausfallsicherheit.

Die Plattformen bieten eine Vielzahl von Schnittstellen: zum Telefonnetz für Sprachkanäle und Signalisierung, sowie zum Abrechnungssystem, zu Datenbanken, Überwachungsmonitoren und Statistiksystemen. Moderne Systeme unterstützen die Applikationsentwicklung und statistische Auswertungen durch grafisch-basierte Tools.

Die Plattform-Hersteller sind in der Regel für die kundenspezifische Integration und die Applikationserstellung verantwortlich. Die Applikationen nutzen auch Basistechnologien zur Sprachverarbeitung, die in der Regel von Drittanbietern stammen. Hierzu zählen insbesondere die Spracherkennung und Sprachsynthese, wobei je nach Plattform und Anforderung unterschiedliche Produkte zur Auswahl stehen.

Bei der Spracherkennung unterscheidet man

- a. Ganzworterkennung weniger Kommandowörter, mit ähnlicher Dialogführung wie bei DTMF („für A sagen Sie ‚eins‘, für B ...“ statt „für A drücken Sie ‚eins‘, für B ...“)
- b. phonembasierte Erkennung großer Wortschätze (VRU führt den Dialog)
- c. phonembasierte Erkennung mit natürlich-sprachlichem Dialog (gemischt-initiative Dialogführung): „Wann fährt morgen ab 14 Uhr ein Zug von Bochum nach Oldenburg?“ Kunde (fühlt sich) bei seinen Spracheingaben nicht eingeschränkt, flache Menüstruktur, daher kürzere Anrufzeiten.

Einige Spracherkennungs-Engines beinhalten zusätzlich eine Sprecherverifizierung, mit der sich die Teilnehmer-Autorisierung, beispielsweise für m-commerce Anwendungen, unterstützen lässt. Sprachsynthesen sind ebenfalls verfügbar, werden in heutigen Applikationen aber nur selten eingesetzt (z.B. in Email-to-Speech).

V. Qualitätsaspekte

Die Qualitätsaspekte einer Sprachdialoganwendung sind vielfältig. Orientiert man sich an den Prinzipien des Dialog-Designs lassen sich folgende Aspekte ausmachen, die alle – mehr oder weniger und ohne Anspruch auf Vollständigkeit – zur Akzeptanz beitragen:

- a. Maximale Erkennungsgenauigkeit:
 - Erkennungsrate
 - Grenzbereiche bzgl. Umgebung, Sprecher oder Kanal
- b. Minimale kognitive Beanspruchung:
 - konzeptionelle Einfachheit
 - Konsistenz
 - globale Befehle wie „Start“, „Hilfe“ etc.
- c. Balance zwischen Effizienz und Klarheit:
 - Prompts: Stil, Wording
 - Word-Spotting
 - Out-of-Grammar-Rate
 - Anpassung an Erst-, Wenig-, Vielnutzer
 - Barge-In
- d. effektive Fehlerbehandlung:
 - Verteilung: falsche Akzeptanz bei In-/Out-of-Grammar
 - falsche Rückweisung bei In-Grammar
 - Fehlermeldungen
 - Transaktionsrate
 - Rückfallmöglichkeiten (z.B. Transfer zum Helpdesk)
- e. Performanz:
 - Antwortzeiten
 - Durchsatz

Die Qualitätsmessung muss in Zukunft wesentlich anwendungsorientierter und unter ‚Real-Life‘-Bedingungen erfolgen. Die Bedeutung vieler Aspekte im Anwendungszusammenhang ist heute noch unklar, ebenso die Methoden zur Messung dieser Aspekte. Kritische Anforderungen an die Tests sind neben der Validität möglichst geringer Aufwand und Reproduzierbarkeit.

VI. Ausblick

Die heute verfügbare *Sprachtechnologie* hat eine Reife erlangt, die eine neue Generation von Bedienschnittstellen und völlig neue Anwendungen erlaubt. Getragen wird diese Entwicklung im wesentlichen durch bessere Ergebnisse der Spracherkennung sowie durch natürlich-sprachliche Dialoge. Aber auch die heutige Synthesequalität bietet breitere Einsatzmöglichkeiten, zunächst offline als Text-to-Speech oder auch online als Concept-to-Speech.

Während heutige VRUs meist als zentrale Netzelemente implementiert sind werden Dienste-Systeme in Zukunft im Netz verteilt, beispielsweise als Komponenten im sog. Intelligenten Netz. Außerdem werden Gateways zu Web, WAP, SMS-Center etc. an Bedeutung gewinnen, da der Grad der Interoperabilität steigen wird. Welche Änderungen sind bei den *Diensten* zu erwarten? Der Sprachzugang wird neben Internet, WAP und SMS bestehen bleiben, denn Sprache ist immer noch die natürlichste Form der Kommunikation und bietet entscheidende Vorteile. Die Dienste werden zunehmend individualisierbar in Form von persönlichen Adressbüchern oder Bookmarks, was die zunehmende Bedienkomplexität teilweise auffangen kann. Insgesamt müssen die Bedienschnittstellen einfacher und effizienter werden, insbesondere wenn man breite Zielgruppen erreichen will.

Zukünftige *Anwendungen*, die von herausragender Bedeutung im Wettbewerb sind und über deren Markterfolg besonders die Bedienbarkeit entscheidet, sind Messaging und Portal Services: Die Voicemail wird zum Unified Messaging System migriert, das den Zugang zu Sprach-, Fax-, Email- und Kurznachrichten über beliebige Kanäle erlaubt wird. Zu den Portalen zählen Zugänge zu mehreren Diensten hinter einer einzigen Servicenummer, Voice Activated Dialling sowie Voice Web. Im Endausbau werden die gesamten Telefoniefunktionen, alle Dienste sowie das Internet über „Voicelinks“ vernetzt und über eine Voice Browser Ebene navigierbar und steuerbar.

VII. Fazit

Sprachbasierte Dienste, welche auf modernen Plattformen die neueste Technologie nutzen, stellen besondere Anforderungen an die Qualitätsmessung. Zunächst sind die wesentlichen Aspekte zu definieren und – stets im Anwendungszusammenhang – bezüglich ihrer Relevanz für die Dienste-Qualität zu bewerten. Entsprechende valide Testmethoden sind zu erarbeiten und möglichst weitgehend zu automatisieren.

Der Endkunde ist wesentlich stärker in den gesamten Entwicklungsprozess einzubeziehen. Dies kann z.B. im Rahmen einer Simulation, eines Prototyping oder eines Friendly-User-Tests geschehen. Bis ein Sprachdienst zur Marktreife gelangt sind jedenfalls mehrere Phasen erforderlich, um die akustischen Modelle, die Grammatiken sowie den Dialog zu optimieren.

Die Auswahl und der Einsatz von Sprachtechnologie sowie das Design der Bedienschnittstelle ist oft nicht optimal möglich, was zu Lasten von Entwicklungszeiten, Kosten und der erzielten Dienste-Qualität geht. Dies lässt sich nur durch die Schaffung eines Know How Pools lösen, in dem Empfehlungen, Methoden, Werkzeuge und Dialogbausteine gesammelt werden. Diesbezüglich wäre eine engere Verzahnung und Austausch mit der Forschung wünschenswert. Dabei kann die Forschung von der Infrastruktur und den Erfahrungen aus dem Produktionsbetrieb im Mobilfunk profitieren. Die Voraussetzungen sind gut: Auf den Dienste-Plattformen wird moderne Sprachtechnologie verfügbar, das Kundenpotential der Netzbetreiber ist groß und der Markt fordert ständig neue und anspruchsvolle Anwendungen. Aber nur durch interdisziplinäre Anstrengungen sowie eine Kooperation zwischen Forschung und Industrie wird es gelingen, dem Endkunden mit vertretbarem Aufwand auch in Zukunft eine optimale Dienste-Qualität gewährleisten zu können.

Referenzen

- [1] INFORMATION WEEK online, „Voice gets reliable“ von Mary E. Thyfault, Ausgabe 22. Februar 1999.
- [2] www.d2mannesmann.de
- [3] „Designing Interactive Speech Systems: From First Ideas To User Testing“, Niels Ole Bersen et al., Springer-Verlag, 1998.