

Technische Basis und Anforderungen an Sprachlösungen

Detlev Artelt

aixvox GmbH
Monheimsallee 22
52062 Aachen
detlev.artelt@aixvox.com

Abstract: Sprachapplikationen bedürfen eine technisch ausgereifte wie auch performante Betriebsgrundlage. Hierzu gibt es neben unterschiedlichen Betriebssystemen auch unterschiedliche Telefonanbindungen und unterschiedliche Softwaresysteme, die eine Sprachlösung erst ermöglichen. Dieser Artikel gibt einen ersten groben Überblick darüber, welche technischen Grundlagen für eine erfolgreiche Sprachapplikation benötigt werden. Fazit: Eine genaue Bedarfsanalyse im Vorfeld erspart Investitionen in die falsche Technologie und sichert somit den langfristigen Erfolg des Sprachsystems nachhaltig.

1 Systemanforderungen

Zu aller erst sollten die Anforderungen an das System identifiziert werden. Ist dies einmal gelungen, lassen sich die nächsten Schritte in der Regel einfacher gehen, denn es macht eine Auswahl der technischen Basis viel einfacher und schließt einige Komponenten ggf. bereits im Vorfeld aus. Folgende Klassifizierungen lassen sich festhalten:

- Größenordnungen SOHO, Enterprise, Carrier, ASP
- Telefonanbindungen Analog, ISDN, VoIP, IN
- Sprachtechnologie Nur Tonwahl, Sprachsteuerung, natürlich sprachlich
- Zielsetzung Inbound, Outbound, beides

2 Basistechnologie und Aufbau

Der Trend geht ganz klar in Richtung verteilter Systeme mit standardisierten Schnittstellen und Komponenten. Dies erleichtert die Skalierbarkeit des Systems und erlaubt es zudem bestimmte Komponenten z.B. die Telefoniekarten mehrfach bzw. übergreifend zu nutzen. Systeme basieren auf Standard Industrie-PC's und entsprechenden Telefoniekarten der gängigen Hersteller. Über diese Karten wird der Zugang zum Telefonnetz entweder per ISDN oder auch per VoIP hergestellt. Zudem erlauben viele der Karten zusätzliche Funktionen im Bereich der Spracherkennung. Bei einer Skalierung des Sprachsystems ist besonders auf die Skalierungsmöglichkeiten der Sprachkarten zu achten.

Auch wenn derzeit noch Vorbehalte gegenüber der Zuverlässigkeit bzw. Ausfallsicherheit derartiger Systeme auf PC Basis bestehen, lassen sich diese über Cluster-Architektur mit entsprechenden Redundanzen in den Griff bekommen. Dies zeigt sich auch bei Herstellern von einst proprietären IVR-Lösungen, die ihren Kunden mittlerweile nur noch eine Software verkaufen und ihnen anbieten, alle notwendigen Hardware-Komponenten mittels einer „shopping list“ selbst zu beschaffen.

3 Sprachanbindung

Die Vorteile die dieser Wandel für den Kunden hat, liegen dabei klar auf der Hand. Eine auf Standard-Komponenten aufbauende Architektur verursacht deutlich geringere Kosten bei der Einführung und Instandhaltung. Aber auch die Hardware-Komponenten an sich sind deutlich billiger und zudem austauschbar geworden. Ein weiterer wichtiger Vorteil dieser modernen Architektur liegt auch in der flexibleren Gestaltung von standortübergreifenden Lösungen, die es ermöglichen Applikationslogik und kritische Kundendaten im Haus zu behalten und den „Gateway“ beispielweise zu einem Voice Application Service Provider zu übertragen oder eine „multi-site“ IVR mit zentraler ASR-Engine aufzubauen. Dies bekommt noch größere Bedeutung durch den Wandel zu IP-basierter Telefonie, der stetig vollzogen wird.

4 Sprachtechnologie

Die Unterschiede zwischen den Anbietern sind indes eher graduell. Die Auswahl sollte hier einen guten Kompromiss zwischen Qualität der Spracherkennung und beanspruchter Rechenleistung darstellen. Weitere wichtige Kriterien sind das Kosten-Nutzen-Verhältnis und die Erfahrungswerte bzw. Fähigkeiten beim „Training“ der Komponenten. Hierbei sollte stets berücksichtigt werden, dass die Wahl des Herstellers oft eine „Entscheidung fürs Leben“ darstellt. Denn insbesondere intensives Training und Tuning eines Spracherkenners ist herstellerspezifisch und geht beim Wechsel des Erkenners verloren. Dies ist umso dramatischer, da sich an dieser Stelle die wahre Qualität und Reife einer Applikation versteckt, die auf den ersten Blick oft nicht sichtbar ist, aber das Ergebnis unzähliger Manntage an Optimierungsarbeit darstellt. Somit sind die Kosten für den Erkennen selbst in vielen Fällen deutlich geringer als die Kosten für dessen Optimierung. Daher kann eine falsche Technologieentscheidung an dieser Stelle hohe Folgekosten mit sich bringen.

5 Aufbau eines Sprachservice

Deutlich schwieriger dagegen wird die Ebene des Dialog-Servers. Hier existieren derzeit einige Softwareprodukte die sich in Bezug auf ihren Funktionsumfang oft nur schwer voneinander abgrenzen lassen bzw. stark überschneiden. So gibt es z.B. Sprachmanagement-Systeme die zugleich eine umfassende Entwicklungsumgebung für das Dialog-Design beinhalten. Auch die ehemals proprietären Systeme sind mittlerweile von umfassenden Applikations-Frameworks umgeben, die Funktionen wie grafisches Dialog-

Design, Reporting oder Management der gesamten Ressourcen ermöglichen (siehe Bild unten). Die umfangreichen Entwicklungsumgebungen können oftmals erst in Bezug auf die weitere zu integrierende Technik ausgewählt werden. Die Entscheidung auf dieser Ebene ist sicherlich sehr stark von individuellen Präferenzen geprägt, allerdings kann eine gute Integrationsfähigkeit in die bestehende Contact Center-Infrastruktur oftmals von größerem Nutzen sein als ein kleiner Handling-Vorteil, der sich durch einen „modernen“ Dialog-Server erzielen lässt.

