„Eyes free - Hands free“ oder „Zeit der Stille“ Ein Demonstrator zur multimodalen Bedienung im Automobil

Klaus Bengler[*], Petra Geutner[+], Bernhard Niedermaier[*], Frank Steffens[+]

[*] BMW Group
[+] Robert Bosch GmbH

Abstract

Das Stichwort “Spracheingabe“ weckt in Zusammenhang mit der Benutzung von Funktionen im Automobil hohe Erwartungen: Gerade für die Bedienung während der Fahrt verspricht „Eyes free - Hands free“ entscheidende Vorteile. Viele Funktionen sind leichter bedienbar und direkt erreichbar. Allerdings stellt die Erkennung gesprochener Sprache in der geräuschvollen Umgebung des Automobils auch hohe Anforderungen an die Gestaltung der **M**ensch-**M**aschine-**I**nteraktion (MMI).
Der von den Firmen BMW und Bosch entwickelte Demonstrator macht multimodale Bedienung im Automobil erlebbar und veranschaulicht die sinnvolle Kombination von Bildschirmdarstellung und Sprachein-/ausgabe sowie das Potential geeigneter Dialogführung bei der Behebung von Erkennungsfehlern. Im Demonstratorfahrzeug können mit Spracheingabe und Lenkradtasten Radio und Navigationssystem bedient werden. Die Ausgaben erfolgen sowohl via Sprachausgabe als auch über den Bordmonitor. Das Zusammenspiel aller Eingaben und Ausgaben wird zentral vom Dialogmanager verwaltet, der wiederum über einen Systemmanager mit den Hardwarekomponenten in Verbindung steht. Der Vortrag geht auf spezifische Fragen ein, die bei der Gestaltung der multimodalen Mensch-Maschine Schnittstelle auftauchen, und diskutiert die Architektur des Demonstrators, wie sie zur Implementierung des multimodalen Systems konzipiert wurde.

Einleitung

Die Gestaltung der Mensch-Maschine-Interaktion im Automobil stellt seit jeher hohe Anforderungen an die verwendeten Technologien und ihren Einsatz in diesem Umfeld. Ziel ist es heute mehr denn je, dem Fahrer eine ständig zunehmende Menge von Funktionen, die mehr oder weniger mit der Fahraufgabe in Zusammenhang stehen, verkehrssicher nutzbar zur Verfügung zu stellen. Damit taucht immer wieder die Frage auf, wie das Zusammenspiel zwischen „Fahraufgabe“ und „Bedienbarkeit“ erfolgreich gesteigert werden kann.
Daneben soll die Bedienung im Automobil natürlich auch komfortabel und möglichst leicht erlernbar sein. Der Einsatz von Spracherkennung verspricht einige der in diesem Zusammenhang immer wieder auftauchenden Problemstellungen zu lösen:

- Bedienung ohne Blickzuwendung
- Beide Hände verbleiben am Lenkrad
- Direkter Zugriff auf eine Vielzahl verschiedener Funktionen
- Intuitive Bedienung

Allerdings bringt die Spracherkennung – insbesonders in einer geräuschvollen Umgebung – auch eine Reihe von Nachteilen mit sich, die ihre Nutzbarkeit für den Fahrer einschränken. (Noyes & Frankish, 1994; Baber 1986):

- Perfekte Erkennung kann nicht vorausgesetzt werden.
- Es kann temporäre Nichtverfügbarkeit in sehr lauter Umgebung auftreten.
- Die Reaktion des Nutzers muß innerhalb einer bestimmten Zeit erfolgen.
- Bedienfehler des Nutzers können mit größeren Konsequenzen verbunden sein.

Quelle: 42. Fachausschusssitzung Anthropotechnik, München, 24.-25.10.2000, "Multimodale Interaktion im Bereich der Fahrzeug- und Prozessführung", DGLR

- Vokabular und Arbeitsweise des Systems müssen erlernt werden.

Die Dialogführung im Automobil hat – im Gegensatz zu anderen Domänen – im Sinn des Bedienungskomforts und der Verkehrssicherheit einigen speziellen Aspekten Rechnung zu tragen (s.a. ISO TC22 SC13 WG8, 2000):

- Der Nutzer erhält sofort und angemessen Rückmeldung auf seine Eingaben.
- Die Rückmeldungen des Systems sind prägnant und gut wahrnehmbar, wobei der Systemzustand jederzeit klar erkennbar ist.
- Graphische Informationen bleiben so lange erhalten wie es erforderlich ist.
- Die Bedienung kann jederzeit und ohne Schaden vom Fahrer unterbrochen werden.
- Der Fahrer wird weder zu zeitkritischen Eingaben aufgefordert, noch wird seine dauernde Aufmerksamkeit bei Eingaben beansprucht.

Die Erfüllung dieser Kriterien formuliert gleichzeitig zu verwirklichende Ziele für den implementierten Demonstrator. Neben der Verbesserung der Erkennungsleistung sind gezielte gestalterische Maßnahmen erforderlich, um die Robustheit und den Komfort des Gesamtsystems zu erhöhen. Dadurch kann der Einsatz von Spracherkennung als Komponente des MMI im Automobil optimiert werden. Folgende Maßnahmen werden getroffen:

- Als Eingabemodalitäten stehen sowohl Sprache als auch die Tasten am Multifunktionslenkrad zur Verfügung. Der Wechsel zwischen den beiden Modalitäten ist jederzeit möglich. Die Kombination der Spracheingabe mit einem „konventionellen“ Eingabeverfahren – in unserem Fall den Lenkradtasten – zu einem multimodalen Eingabemedium steigert zwar nicht die Erkennungsleistung der Einzelmodalität, verspricht aber eine Verbesserung des Gesamtsystems.

- Die Bedienung kann jederzeit unterbrochen und wieder aufgenommen werden. Dialog und Bildschirmdarstellung verweilen im aktuellen Zustand, der Spracherkenner schaltet ab.

- Durch die Definition von Dialogen und Subdialogen wird die Anzahl der Eingaben zum Start von Sprachdialogen (PTT) minimiert.

- Da die Rückmeldung des Systems sowohl optisch als auch akustisch erfolgt, bleibt es dem Nutzer überlassen, wie er sich über den aktuellen Systemzustand informiert.

- Systemfeedback erfolgt in Abhängigkeit der Erkennungswahrscheinlichkeit und des Dialogzustandes. Unsichere Erkennungen lösen Konkretisierungsfragen und explizite Prompts aus; sichere Erkennungen münden in implizite Prompts des Systems. Das wohlbekannte „Bitte wiederholen“ ist weder komfortabel noch dient es in den meisten Fällen der Problemlösung. Durch die Kombination verschiedener Strategien wird die Fehlerbehebung erleichtert: Wiederholung, Auswahl aus möglichen Alternativen, Eingabe per Taste.

- Der Benutzer kann Sprachausgaben durch Spracheingaben unterbrechen („barge-in“), wodurch er jederzeit die Kontrolle über den Dialogablauf behält. Gleichzeitig kann er damit die Dauer der Bedienung merklich verkürzen.

- Um den Lernprozeß zu unterstützen, sind für die verwendeten Kommandoworte Synonyme definiert. Für das gesamte System gilt der Grundsatz: „Speak what you see“. Alle angezeigten Befehle können als Sprachkommandos gesprochen werden.

- Die Erkennungsleistung wird durch verschiedene Ansätze in der Dialogführung gesteigert. Hierzu zählen die Nutzung von Zusatzinformationen wie Sonderziele, Adress- und Zielspeicher. Darüberhinaus dienen auch Dialogelemente wie Buchstabieren und Anwahl des ersten Buchstabens per Taste bei einem großen Vokabularumfang wie der Navigationszieleingabe dazu, den Suchraum möglichst einfach einzugrenzen.

Der Bedarf an entwicklungsbegleitenden Prototypen für MMI-Konzepte steigt ständig. Diese Tatsache erfordert gerade im Bereich der Mensch-Maschine-Interaktion immer größere Anstrengungen - vor allem vor dem Hintergrund, daß im Fall des multimodalen MMI unterschiedlichste Technologien integriert werden müssen, um ein „look and feel“ zu ermöglichen.

Funktionalität und Architektur des Demonstrators

Der Funktionsumfang des Demonstrators umfaßt die Bedienung des Radios und eines Navigationssystems. Per Spracheingabe kann zum Beispiel aus der Liste der aktuell empfangbaren Sendernamen ein Radiosender ausgewählt werden. Die Liste der „ansprechbaren“ Sendernamen wird mittels dynamisch generierter Wortschätze zur Laufzeit erzeugt. Das System ist somit in der Lage, neue Sendernamen jederzeit in die Liste aufzunehmen. Der Vorteil dieser Methode wird sofort deutlich, wenn Sender über ihren Namen eingestellt werden können und nicht mehr länger über den Umweg einer Ziffer („1 für Sender x, 2 für...“) aus einer Liste gewählt werden müssen. Daneben sind mit RDS und Verkehrsfunk einige einfache Funktionen des Radios implementiert.
Den Schwerpunkt in der Navigationsanwendung bildet ein fehlerrobuster und komfortabler Dialog zur multimodalen Eingabe von Navigationszielen auf verschiedensten Wegen. Neben der bekannten Eingabe über Ortsname und Strasse stehen auch verschiedene Zielspeicher und Sonderziele zur Verfügung. Hierzu können die 300 größten deutschen Städte mit allen zugehörigen Straßen eingegeben werden.

Hardwarearchitektur

Bei der Auswahl der Komponenten wurde darauf Wert gelegt, eine stabile Funktionsweise mit einem gewissen Grad an Flexibilität zu erreichen. Der Funktionsumfang sollte nicht simuliert sondern unter Verwendung von Serienkomponenten implementiert werden.
Es wurden ein kommerziell verfügbares Radio und Navigationssystem der Firma Bosch über RS232 bzw CAN-Interface an einen WindowsNT Rechner angebunden. Die Ausgaben des Systems erfolgen sowohl via Sprachausgabe als auch über den serienmäßigen Bordmonitor des Demonstratorfahrzeugs. Für die manuelle Bedienung wird das BMW Multifunktionslenkrad in seiner serienmäßigen Ausführung eingesetzt.

Abb 1. Interieur des Demonstrators mit Multifunktionslenkrad und Bordmonitor

Softwarearchitektur

Für die Teilaufgaben Dialogmanagement, Hardware-/ Systemmanagement, Spracherkennung, Sprachausgabe und Visualisierung (GUI) sind jeweils eigene Softwaremodule implementiert.
Die Spracherkennung erfolgt über die kommerziell verfügbare Software ASR 1600 SDK v3.22 der Firma Lernout & Hauspie und die Sprachsynthese mittels TTS Speak & Win Version 3.0 Release 1.0 der Firma ELAN. Mit Hilfe des Tools Altia von I-Logix wird die graphische Bedienoberfläche realisiert.
Die zentrale Verwaltung aller Ein- und Ausgaben übernimmt das Modul „Dialogmanager“, das unter Statemate Magnum 1.3.1 (STMM) implementiert wurde. Dieser Dialogmanager steht über eine Systemmanagementsoftware mit den Hardwarekomponenten für Funktion und Lenkradbedienung in Verbindung und realisiert über die MMI-Module TextToSpeech (TTS), Spracherkennung (ASR) und Visualisation die Interaktion mit dem Nutzer.
Alle Softwarekomponenten sind auf einem PC unter Windows-NT 4.0 integriert.

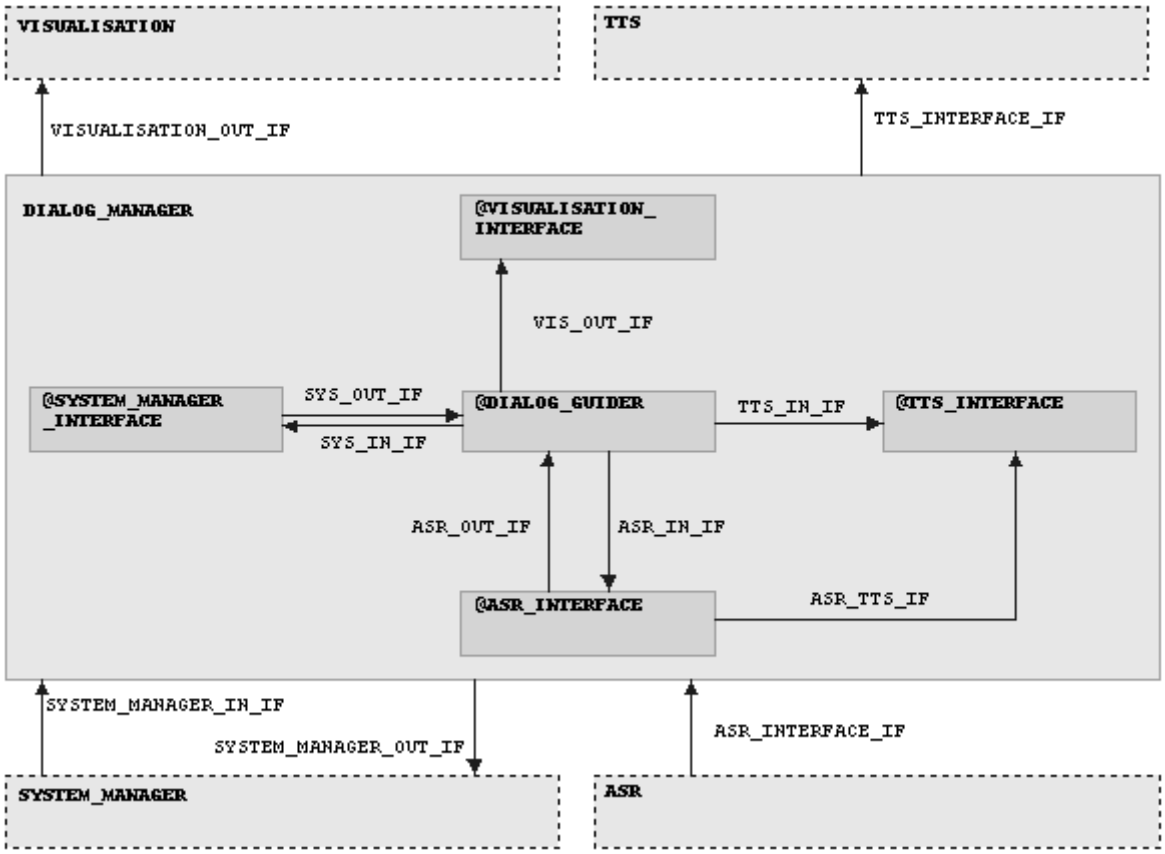

Abb 2. Gesamtsystem mit Schnittstellen zu den übrigen Modulen

Universelle Schnittstellen

Die Definition aller Schnittstellen ist so abstrakt wie möglich gehalten. Die einzelnen Module tauschen lediglich die benötigten Informationen untereinander aus. Diese starke Kapselung der Funktionalitäten erlaubt den Austausch einzelner Schnittstellenmodule, z.B. des TextToSpeech-Systems, ohne die Dialogführung und andere Module anpassen zu müssen.

Dialogmanagement

Der Dialogmanager als Modul greift nicht direkt auf Ein- und Ausgabedaten zu. Alle Dialogdaten, Vokabularien, Aus- und Eingabetexte werden konfigurierbar in den einzelnen Subkomponenten erzeugt und verwaltet. Die Kommunikation zwischen den Teilmodulen erfolgt auf einer abstrakten Ebene über semantische Tokens.

Die zentrale Komponente des Dialogmanagements stellt der DIALOG_GUIDER dar. Als hierarchischer Zustandsautomat verwaltet er die einzelnen Dialogzustände, synchronisiert die Subkomponenten und führt Zustandsübergänge durch. Auf diese Weise wird ein konsistentes Verhalten der Subkomponenten garantiert.

Dialogsteuerung

Während andere Ansätze z.T. auf agentenbasierten Architekturen und Heuristiken beruhen (vgl. Nigay und Coutaz, 1995) wird der gesamte Dialog streng deterministisch mittels Statecharts gesteuert, wobei für jeden Dialogzustand ein eigenes Statechart angelegt wurde. Beim Eintreffen eines entsprechenden semantischen Tokens aus einer Subkomponente führt der DIALOG_GUIDER einen Zustandsübergang aus, der wiederum mittels eines entsprechenden Kommandos an die Subkomponeten gemeldet wird.

Je nach ihrem Gültigkeitsbereich sind lokale und globale Kommandos zu unterscheiden. Lokale Kommandos sind nur im jeweiligen Dialogzustand gültig und können dort kontextbezogen verwendet werden. Ein Beispiel dafür sind Eingabebestätigungen.

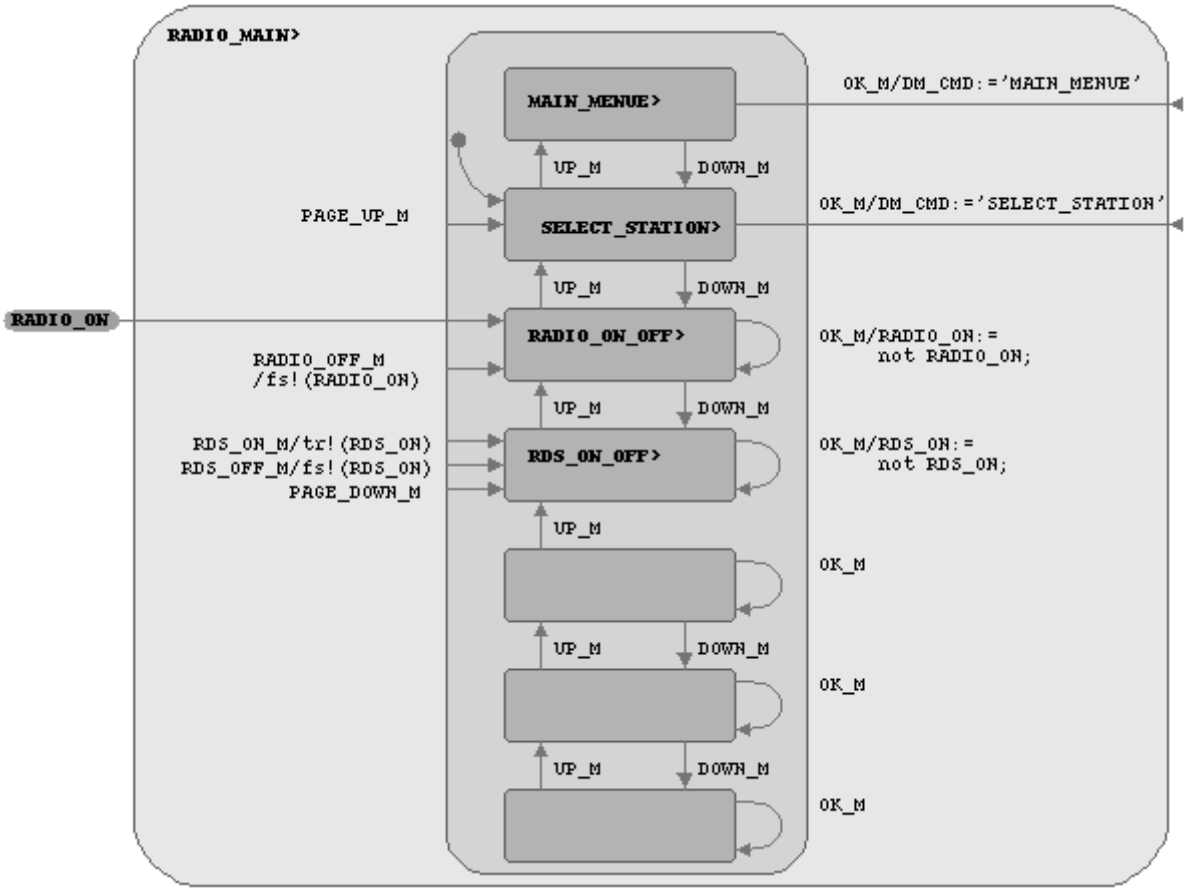

Abb 3. Beispiel: RADIO_MAIN - Statechart

Mit Hilfe globaler Kommandos können bestimmte Dialogzustände unabhängig vom aktuellen Dialogzustand aus erreicht werden. Sie sind daher nicht kontextbezogen einsetzbar. Insbesondere bei der Funktionsauswahl per Sprachkommando und zum Erreichen eines für den Benutzer transparenten Dialogzustandes stellen globale Kommandos ein sehr effizientes Instrument dar.
Das Beispiel der Zustände „CITY_MENUES“ und „NAVIGATION_MENUES“ zeigt eine solche hierarchisch aufgebaut Struktur. Bei globalen Kommandos wird der gesamte Menüzustand verlassen und dann in den gewünschten Zweig gesprungen.

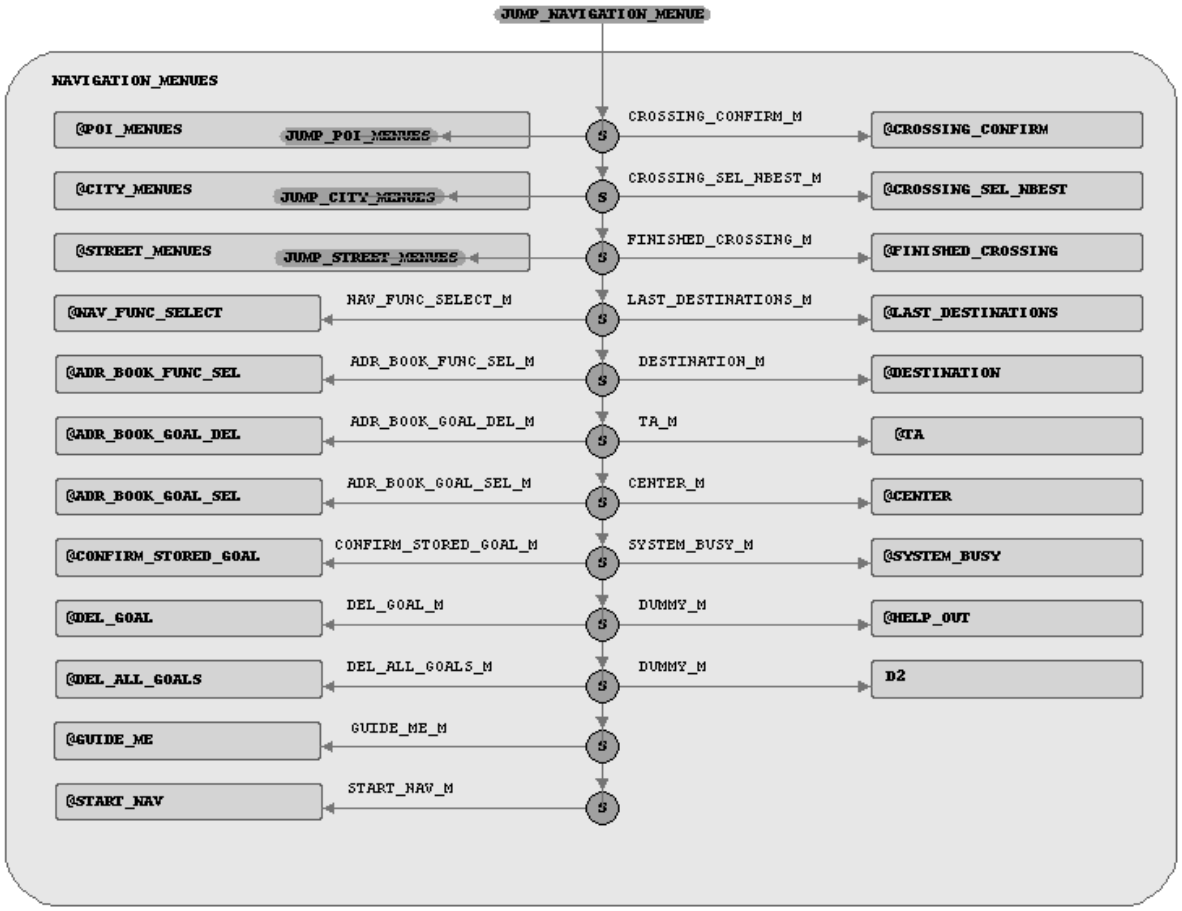

Abb 4. Implementierung globaler Kommandos

Subkomponenten

Die einzelnen Subkomponenten Sprachein-/ausgabe, Systemmanagement und Visualisierung sind als eigenständige C-Libraries implementiert und kapseln so die funktionalen Details. STMM bietet ein Interface, über das die entsprechenden Subroutinen aufgerufen werden können. Beim Kompiliervorgang erfolgt dann die feste Bindung der einzelnen Bibiliotheken an das Statemate-Projekt. Um gegenseitige Blockaden zu verhindern, wird jede Subkomponente in einem eigenen Thread gestartet. Insbesondere die Bibliothek für den Spracherkenner gestaltet sich aufwendig, da Dienstfunktionen für das Erzeugen dynamischer Vokabularien also auch für die Handhabung großer vorkompilierter Vokabularien notwendig waren.

Zusammenfassung

Der vorgestellte Demonstrator zeigt eine Hard- und Softwarearchitektur für ein multimodales Bedienkonzept. Bei der Erstellung des Systems wurde hauptsächlich auf die Anforderung geachtet, die an den Demonstrator als Tool der Produktentwicklung gestellt wird: Flexibilität auf unterschiedlichen Ebenen. Die aktuelle Implementierung erlaubt daher ohne größeren Aufwand einige der dargestellten Inhalte zu ändern:

- Sprache/Nationalität der Sprachein-/ausgabe
- Inhalte der Sprachausgaben
- Funktionsweise der Bedienelemente am Lenkrad
- Zusammenstellung und Funktionsweise der Kommandowörter
- Austausch der verwendeten Softwaremodule

Ausschlaggebend ist hierbei, daß diese Flexibilität vor dem Hintergrund eines stabilen und konsistenten Dialogverhaltens zur Verfügung steht, das seinerseits in einem gekapselten Softwaremodul implementiert wurde.

Das MMI Konzept greift einige der Belange auf, die für komfortable Bedienung im Fahrzeug relevant sind, und zeigt, daß Multimodalität den Einsatz von Spracheingabe im Fahrzeug verbessert. In Versuchen muß nun evaluiert werden, in welchem Ausmaß die Kombination der Spracheingabe mit Tastenbedienung den Bedienkomfort steigert und die graphische Rückmeldung des Systems die Erlernbarkeit erhöht. Von Interesse ist auch wie sich unterschiedliche Systemrückmeldungen und Dialoge in Abhängigkeit der Erkennungswahrscheinlichkeit auf die Erlernbarkeit auswirken.

Für die Weiterarbeit zeichnet sich ab, daß für die schnelle Erstellung hochwertiger MMI Konzepte zur Bedienung komplexer Funktionen im Automobil bisher noch die geeigneten Entwicklungswerkzeuge fehlen. Wie schon von Nigay und Coutaz (1995) beklagt, wurde auch im hier vorgestellten Projekt ein Großteil der Resourcen dafür aufgewendet, die Einzelkomponenten zu integrieren. Darüberhinaus zeichnet sich ab, daß für die Evaluation multimodaler MMI-Konzepte noch geeignete Methodiken fehlen.

Literatur

Baber, C. (1996). Automatic Speech Recognition in Adverse Environments. In Human Factors 38,1 (pp. 142-155).

ISO TC22 SC13 WG8 (2000). „Dialogue Management Compliance". Draft version published at the Task Force Meeting 2000.

Nigay, L. & Coutaz, J. (1995). A Generic Platform for Addressing the Multimodal Challenge. Proceedings of the CHI '95.

Noyes, J.M. & Frankish, C.R. (1994). Errors and Error Correction in Automatic Speech Recognition Systems. In Ergonomics 37,11 (pp. 1943-1957).