

Komfortable Sonderzieleingabe per Sprache im Automobil

Kerstin Müller, Stefan W. Hamerich

Harman/Becker Automotive Systems GmbH – Speech Dialog Systems – Ulm

Zusammenfassung

In diesem Beitrag werden die Sprachbedienung und ihr technischer Hintergrund im Automobil erläutert. Insbesondere wird der momentane Stand der Sonderzieleingabe vorgestellt und analysiert, darauf aufbauend werden zwei neue komfortablere Ansätze zur Spracheingabe von Sonderzielen vorgestellt.

1 Einleitung

Mit dem Einzug der Sprachbediensysteme in Mittel- und Oberklassefahrzeugen eröffnen sich neue Möglichkeiten der Gerätebedienung. Die haptische Bedienung wird mehr und mehr durch die Steuerung per Sprache ergänzt.

Dieses Papier ist wie folgt aufgebaut: Im nächsten Abschnitt erfolgt die Einführung in die automobilen Sprachbedienung. Anschließend werden in Abschnitt drei die Sonderzieleingabe in heutigen Systemen und die neuen Ansätze dargestellt. Zum Abschluss folgen eine Zusammenfassung und ein Ausblick.

2 Sprachbedienung im Automobil

Sprachbediensysteme im Automobil sind seit 1996 auf dem Markt erhältlich (Heisterkamp 2001). Das erste System erlaubte mittels 30 sprecherunabhängiger Kommandos die Bedienung des fest eingebauten Autotelefons. So konnte z.B. eine Telefonnummer komfortabel während der Fahrt per Sprache gewählt werden.

Heute sind Sprachbediensysteme zu einem wichtigen Bestandteil der Zubehörlisten von fast allen Mittel- und Oberklassewagen geworden. Systeme von Harman/Becker als einem der führenden Hersteller sind aktuell in Fahrzeugen von Audi, BMW, Lancia, Maybach, Mercedes-Benz, Porsche, Rolls-Royce und weiteren namhaften Herstellern verbaut. Damit kann die

Sprachbedienung im Automobil als eine Erfolgsgeschichte der Sprachverarbeitung bezeichnet werden (Hanrieder 2004).

Neben der Kontrolle des Autotelefons erlauben Sprachbediensysteme inzwischen auch die Bedienung von Radio, Klimaanlage und Medienspieler. Seit einiger Zeit ist auch die Sprachbedienung des Navigationssystems verfügbar. Ein besonderes Feature ist dabei die komfortable Eingabe eines Navigationsziels per Sprache. Damit sind jetzt z.B. in Deutschland fast 80.000 Ortsnamen sprechbar. Die bereits angesprochene große Datenmenge macht die komfortable Navigationszieleingabe im Kfz sehr aufwendig. Daher sind die meisten existierenden Systeme noch auf das Buchstabieren der Ortsnamen angewiesen.

Der sprachliche Zugriff auf größere Datenmengen ist derzeit noch im Entwicklungsstadium. Ein Thema dabei ist auch die direkte Ansteuerung von MP3-Sammlungen per Künstler oder Titel, wie z.B. „*Spiele ,99 Luftballons*“, vgl. (Wang et al. 2005). Des Weiteren wird an der Spracheingabe von Sonderzielen gearbeitet, welches das Hauptthema dieses Beitrags ist.

Die Sprachsteuerung im Automobil unterscheidet sich in wesentlichen Aspekten von Diktiersystemen auf dem PC oder telefonischen Dialogsystemen. Die Grundlagen der Spracherkennung sind allerdings die gleichen, vgl. (Jelinek 1998). Der wesentliche Unterschied liegt in der Integration der Sprachbediensysteme, die üblicherweise ab Werk installiert werden. Während zur Spracherkennung auf einem PC mehrere hundert Megabytes heutzutage Standard sind, muss ein eingebettetes System im Automobil in der Regel mit signifikant weniger Speicher auskommen. Diese Diskrepanz begründet sich mit den Anforderungen der Automobilindustrie, welche neben der Einsetzbarkeit unter unterschiedlichen Klimabedingungen auch eine einwandfreie Funktion bei Erschütterungen während der Fahrt fordert. Näheres siehe bei (Hamerich 2005).

3 Sonderzieleingabe

Ein Sonderziel ist ein Ort von z.B. kulturellem oder touristischem Interesse, mithin ein für die Allgemeinheit interessantes Ziel, wie z.B. das *Ulmer Münster*, die *nächste Tankstelle* oder das *nächste Krankenhaus*. Alle Ziele, die außer der Adresse auch über einen Eigennamen eindeutig identifizierbar sind, können als Sonderziele (engl. Points of Interest – POI) bezeichnet werden.

Die Sonderzieleingabe im Fahrzeug dient dazu, verschiedene Ziele über deren Eigennamen einzugeben, ohne dass der Benutzer die genaue Adresse kennen muss. Bei der haptischen Bedienung wird der Benutzer durch mehrere große strukturierte Listen geführt. Der Vorteil der Sprachbedienung ist die Möglichkeit, Menüebenen überspringen zu können.

In aktuellen Systemen wird üblicherweise zuerst der Zielort vom Benutzer abgefragt. Anschließend wird der Nutzer durch eine meist einstufige Kategorienstruktur geführt, in welche die Sonderziele einsortiert sind. Dabei sind in vielen Fällen die Sonderziele nur einer Kategorie zugeordnet. Letzteres kann die Suche nach einem bestimmten POI sehr erschweren, da der Großteil der Sonderziele nicht nur einer Kategorie zugeordnet werden kann. Dies kann

dazu führen, dass z.B. das *Ulmer Münster* nicht unter der Kategorie *Gotteshäuser* zu finden ist, sondern nur unter *Sehenswürdigkeiten*.

Um die prototypische Umsetzung der neuen Ansätze zu testen, wurden diese in einer PC-Simulation implementiert, die auf der gleichen Dialogentwicklungsumgebung (vgl. dazu (Hamerich und Hanrieder 2004)) beruht, mit denen auch die Produkte implementiert werden. Die Navigations-Datenbasis wurde für die Prototypen aus Zeitgründen auf die Stadt Ulm und deren Umgebung reduziert.

3.1 Kategorienansatz

Der weiterentwickelte Kategorienansatz basiert nach der Eingabe des Zielortes bzw. des Umkreises auf einer mehrstufigen Kategorienstruktur sowie einer multiplen Zuordnung der Sonderziele zu mehreren passenden Kategorien.

Dazu wurde im Dezember 2006 eine Webumfrage mit über 400 Teilnehmern (darunter knapp 350 Nutzer von Navigationsgeräten) durchgeführt, welche eine Zuordnung von Sonderzielen zu Kategorien bzw. Unterkategorien zu Oberkategorien ergab. Die intuitive Kategorieneinteilung ist für die Benutzerzufriedenheit sehr wichtig, da nur so gewährleistet werden kann, dass Benutzer ihr gewünschtes Sonderziel im ersten Versuch finden.

Da sich Benutzer erwiesenermaßen in bis zu drei Ebenen gut zurechtfinden können (Salmen 2002), wurde dieser Ansatz für drei Ebenen spezifiziert. Damit werden die Listen für den Anwender deutlich kürzer, was zu einer komfortableren Listennavigation führt.

Die Kategorien wurden in der obersten Ebene in allgemeine Begriffe wie z.B. Freizeit, Verkehr, Tanken, Touristisches etc. gefasst, um eine erste grobe Sortierung zu erreichen. In der darunterliegenden Ebene wurden konkretere Begriffe verwendet, sodass die Listen überschaubarer werden. Zudem bietet dieser Ansatz durch das eingeschränkte Vokabular sehr gute Erkennraten.

Um für erfahrene Benutzer die Bedienung weiter zu beschleunigen, können auch die nicht sichtbaren Ebenen gesprochen und erkannt werden. Diese sog. *Shortcuts* stellen einen großen Vorteil der Sprache gegenüber der haptischen Bedienung dar, da mehrere Menüebenen einfach übersprungen werden können.

3.2 Stichwortansatz

Im Stichwortansatz wird dem Benutzer völlig freigestellt, welche Informationen er wann und in welcher Reihenfolge dem System übermitteln möchte. Das System erkennt die Äußerung des Benutzers und überprüft sie auf die Komponenten Stadt, Kategorie und Item. Bei Fehlen einer eindeutigen Komponente fragt das System diese beim Benutzer nach. So kann der Nutzer im Idealfall mit einer Spracheingabe ein Sonderziel eingeben.

Problematisch wird es bei Äußerungen wie *Postamt Neustadt*, die aufgrund der Mehrdeutigkeit des Namens nicht einer eindeutigen Stadt zuzuordnen sind. Hier hilft bei der Disambigu-

ierung der 31 Neustadts in Deutschland die Abfrage der Postleitzahl oder des Namenszusatzes, um dem Benutzer einen Anhaltspunkt zu geben, welches Neustadt gemeint ist.

Einzelne Äußerungen, wie nur die Stadt oder nur eine allgemeine Kategorie zu sprechen, sind in diesem Ansatz auch möglich, der Benutzer wird in diesem Fall durch die aus der Webumfrage hervorgegangene Kategorienstruktur geleitet.

Dieser Ansatz ist sehr benutzerfreundlich, da er fast ohne Menüstruktur auskommt und sowohl von Anfängern als auch von erfahrenen Benutzern sehr leicht zu verwenden ist. Er verlangt allerdings einen großen Erkennerschatz, der für große Länder noch nicht im Endsystem zur Verfügung steht.

4 Zusammenfassung und Ausblick

In diesem Papier wurde die sprachliche Eingabe von Sonderzielen in ein Navigationssystem im Automobil beschrieben. Es wurde dargestellt, dass mittels der Spracheingabe die Sonderzielnavigation komfortabler und benutzerfreundlicher gestaltet werden kann, dies aber momentan aus technischen Gründen noch nicht im Endsystem verfügbar gemacht werden kann.

Die aktuell laufende Evaluation der Ansätze wird zeigen, wie die Benutzer auf die neue Eingabemöglichkeit reagieren, und soll das Verbesserungspotenzial dieser beiden theoretischen Ansätze aufzeigen.

Nach einigen weiteren Iterationen und bei Nutzung einer leistungsfähigeren Hardware können seriennahe Tests mit größeren Datenmengen durchgeführt werden. Anschließend erst kann diese Funktionalität für Endkunden aufbereitet werden. Damit ist mit einer Erlebbarkeit dieser Funktion für Endkunden erst in einigen Jahren zu rechnen.

Literatur

- Hamerich, S. W. (2005): Speech Dialogue Systems for Cars – an Overview. SDV – Sprache und Datenverarbeitung, Bd. 29, Nr. 2, Seiten 107–118.
- Hamerich, S. W., Hanrieder, G. (2004): Modelling Generic Dialog Applications for Embedded Systems. In: Tagungsband ICSLP. Jeju, Korea, Seiten 237–240.
- Hanrieder, G. (2004): Sprachbedienung im KFZ – Eine Erfolgsgeschichte. In: Tagungsband 'Informatik 2004'. Ulm, 2004, Seiten 220–224.
- Heisterkamp, P. (2001): Linguatronic – Product-Level Speech System for Mercedes-Benz Cars. In: Tagungsband HLT. San Diego, USA, Seiten 1–2.
- Jelinek, F. (1998): Statistical Methods for Speech Recognition. MIT Press, Cambridge, USA.
- Salmen, A. (2002): Multimodale Menüausgabe im Fahrzeug, Philosophische Fakultät IV der Universität Regensburg. Erschienen im Verlag Herbert Utz, Dissertation.
- Wang, Y.-F. H., Hamerich, S. W., Hennecke, M. E. und Schubert, V. M. (2005): Speech-controlled Media File Selection on Embedded Systems. In: Tagungsband SIGdial. Lissabon, Portugal, Seiten 217–221.