Evaluierung der Erkennungssicherheit von sprachbedienten Systemen im Kfz

Martin Herrenkind[1], Dr. Gudrun Klasmeyer[2], Karl Kreft[1]

[1] *IAV GmbH, 38518 Gifhorn, Deutschland, Email: martin.herrenkind@iav.de, karl.kreft@iav.de*

[2] *IAV GmbH, 10587 Berlin, Deutschland, Email: gudrun.klasmeyer@iav.de*

Einleitung

Da die Komplexität von Systemen mit Infotainment- und integrierter Komfortfunktionalität im Kraftfahrzeug in den vergangenen Jahren rasant zugenommen hat, bemühen sich Automobilhersteller um *Human Machine* Interfaces, die diese schier unüberschaubare Vielfalt von Funktionen für den Fahrer beherrschbar machen, ohne seine Aufmerksamkeit übermäßig vom Straßenverkehr abzulenken. Neben haptischen Eingabemöglichkeiten werden verstärkt akustische Ein- und Ausgaben realisiert.

Die Integration von sprachbedienten Systemen in Kraftfahrzeuge ist ein vielschichtiger Prozess, an dem mehrere Partner beteiligt sind. Zwar hat der Lieferant des Sprachbediensystems vor Beginn des Entwicklungsprojektes die grundsätzliche Eignung des verwendeten Spracherkenners nachzuweisen, jedoch bleibt für den Fahrzeughersteller die Notwendigkeit, das Zusammenspiel mit den sprachbedienten Komponenten sowie die Bedienqualität im jeweiligen Zielfahrzeug zu verifizieren und abzusichern.

Als Gütemaß für den Spracherkenner wird dabei eine Erkennungsrate in Prozent ermittelt. Dieses Gütemaß kann am Ende der Entwicklungsphase als pauschales Urteil für das komplette Sprachbediensystem ermittelt werden. Jedoch findet die Qualitätssicherung im Idealfall nicht erst zum Ende der Entwicklungsphase statt, sondern bereits entwicklungsbegleitend, da sprachbediente Systeme für mehrere Sprachen parallel entwickelt werden und das eventuell nötige Nachbessern akustischer Modelle im Spracherkenner mit einem höheren Zeitaufwand einhergeht.

Definition eines Validierungsprozesses

Neben den verschiedenen Sprachen gibt es andere Variablen, die bei der Evaluierung eines Sprachbediensystems im Kraftfahrzeug zu berücksichtigen sind. Deshalb soll hier ein Validierungsprozess definiert werden, der es erlaubt,

- reproduzierbare und generalisierbare Aussagen über die Güte eines Sprachbediensystems im praktischen Einsatz zu machen,
- woraus sich schon während der Entwicklungsphase Hinweise ableiten lassen, wo systematische Verbesserungen ansetzen könnten und
- mit deren Hilfe sich die Wirkung getroffener Maßnahmen eindeutig messen und belegen lässt.

Selbstverständlich ist wie bei allen Prozessen in der Automobilindustrie eine Kostenminimierung zu berücksichtigen, so dass

- maximale statistische Aussagekraft unter Einsatz von minimalem Aufwand erzielt werden muss.

1. "Statische" Sprechermerkmale

In der Praxis wird die Bewertung eines Spracherkenners häufig von wenigen Einzelpersonen vorgenommen, die im Rahmen einer Funktionserprobung für ein komplettes System (z. B. Navigation) die Sprachbedienung dieses Systems mit zu bewerten haben. Die Einzeltester können durchaus eine prozentuale Angabe der von ihnen untersuchten "Erkennungsrate" machen.

Die Aussagekraft dieser Messung ist jedoch nur sehr gering, da die im Spracherkenner hinterlegten akustischen Modelle mit einigen wenigen Sprecher, die die Bandbreite akustischer Merkmale der zukünftigen Systemnutzer gar nicht abdecken, nicht ausreichend getestet werden können. Der geübte Tester lernt darüber hinaus während der Benutzung des Systems seine Stimme so zu verändern, dass die Erkennungsrate maximal wird.

Diese Vorgehensweise beinhaltet ein hohes Risikopotential für den Automobilhersteller, da diese "Messung" u.U. Richtungsentscheidungen im Projekt auslöst, die keine ausreichende Grundlage besitzen. Um dieses Risikopotential zu minimieren, muss zunächst eine Analyse der Merkmale des Systemnutzers (Kunden) durchgeführt werden:

Sprechermerkmale, die neben der Muttersprache in direktem Zusammenhang mit den akustischen Merkmalen der Sprachbefehle stehen, mit denen das System bedient wird, sind:

- Alter
- Geschlecht
- dialektale Färbung (regional unterschiedliche Aussprache bei gleichem Vokabular sowie regionale Unterschiede im Vokabular selbst).

Unter Berücksichtigung dieser Faktoren kann eine repräsentative Stichprobe aus dem avisierten Kundenkreis gezogen werden.

2.Geräuschkulisse und "dynamische"Sprechermerkmale

Ein weiterer Aspekt bei der Testplanung ist die Berücksichtigung von Umwelteinflüssen in Form von Hintergrundgeräusch in der Fahrzeugkabine. Das Geräusch hat je nach Pegel und Zusammensetzung nicht nur direkten Einfluss auf den Spracherkenner, sondern auch Einfluss auch die Artikulation des Sprechers und damit indirekt auf den Spracherkenner. Das Hintergrundgeräusch bestimmt so die Herausbildung von dynamischen Sprechermerkmalen (Lombardeffekt), die neben den zuvor genannten statischen Sprechermerkmalen zu berücksichtigen sind.

Merkmale zur Charakterisierung des Hintergrundgeräusches sind:

- Fahrzeugausführung (Limousine, Kombi, Coupé,...)
- Fahrzeuginnenausstattung
- Motorisierung
- Fahrgeschwindigkeit
- Betriebsart von Lüftung / Klimatisierung
- Fahrbahnbeschaffenheit

- Wetter (z.B.: Regen, Hagel,....)
- andere externe Umgebungsgeräusche
- Öffnen der Fenster
- Betrieb von Nebenaggregaten (z.B.: Scheibenwischer, Sitzverstellung,....)

Für das zu betrachtende Fahrzeug ist daher ein Katalog von Fahrprofilen zu erstellen und im Test zu berücksichtigen, die der typischen Nutzung des Fahrzeugs, bzw. des Spracherkenners in der jeweiligen Fahrsituation entsprechen.

3. Vokabular

Als dritter Testfaktor kommt die Auswahl des im Spracherkenner berücksichtigten Vokabulars hinzu. Handelt es sich bei dem betrachteten sprachbedienten System beispielsweise um eine Freisprecheinrichtung mit einigen Dutzend Systembefehlen, so stellt es keinen großen Aufwand dar, diesen Wortschatz im Test komplett abzudecken.

Moderne DVD basierte Navigationssysteme können jedoch bereits einige hundert Systembefehle und hunderttausende Städte- und Straßennamen (z.B. europaweit) unterscheiden. Es scheint nicht realistisch, diesen Wortschatz im Rahmen eines Tests abzudecken. Die Zusammenstellung einer geeigneten Stichprobe ist daher für das Testergebnis entscheidend. Die Auswahl der verwendeten Namen und Bezeichnungen muss so erfolgen, dass sowohl nach statistischen als auch nach phonetischen Merkmalen eine repräsentative Stichprobe des gesamten Vokabulars gezogen wird. Die Generalisierbarkeit des Messergebnisses ist nur dann gewährleistet, wenn innerhalb der Stichprobe alle typischen Merkmale des zu beurteilenden Vokabulars abgedeckt werden.

Zu einem repräsentativen Testset gehören an dieser Stelle:

- häufig angefahrene Ziele
- große Orte
- häufige Straßennamen

aber auch:

- Orte und Straßen mit sehr kurzen Namen
- Orte und Straßen mit regionaltypischen Namen
- ungewöhnlich konstruierte Straßennamen (wie "*B Avenue" in New York, "Strasse 47" in Berlin,...*)[1]

Reproduzierbarkeit von Ergebnissen

Der Testprozess muss so gestaltet werden, dass die Wiederholung einer Messung zum gleichen Ergebnis führt. Im Fall einer Abweichung des Ergebnisses darf nur eine Modifikation am untersuchten System die Ursache sein, andernfalls sind Erfolg bzw. Misserfolg einer Änderung oder Weiterentwicklung am System nicht nachweisbar.

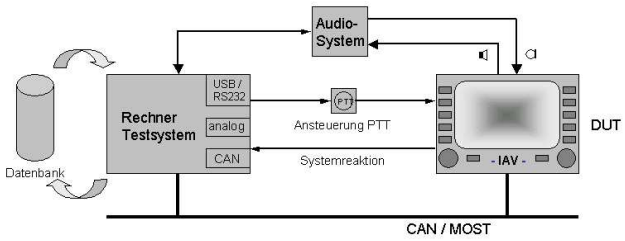
Zur exakten Wiederholung eines Tests können reale Sprecher nur bedingt herangezogen werden, da ihre Artikulation abhängig von der Tagesform ist. Außerdem ist es wenig ratsam, die Testwiederholung von der Verfügbarkeit bestimmter Sprecher abhängig zumachen. Nicht zuletzt stellt die wiederholte Verwendung realer Sprecher einen nicht zu vernachlässigenden Kostenfaktor dar. Lösung ist an dieser Stelle die Verwendung von Sprachsamples, die im reflexionsarmen Raum aufgezeichnet wurden, und für den jeweiligen Test unter Berücksichtigung der zuvor diskutierten Testfaktoren aus einer Datenbank pegelrichtig in das zu beurteilende Zielfahrzeug eingespielt werden. (Siehe dazu auch [2], S.2, Abschnitt A.)

Ähnliches gilt für das Hintergrundgeräusch. Zwar sind die festgelegten Fahrprofile (s.o.) zumindest auf speziellen Testgeländen reproduzierbar, die Umweltbedingungen können jedoch stark voneinander abweichen (Risikofaktoren: Wetter, externe Umgebungsgeräusche). Die Messung sollte daher möglichst nicht im realen Fahrbetrieb, sondern unter Laborbedingungen durchgeführt werden. Nur so lassen sich die zahlreichen variablen Faktoren in feste Größen überführen. (Siehe dazu auch [2], S.2, Abschnitt B.)

Durch die Simulation der Fahrzeugumgebung wird aus Sicht des zu bewertenden Systems das Mikrofonsignal vollständig nachgebildet und die Reproduzierbarkeit der Messung ist gewährleistet.

Weitere Aufwandsminimierung

Eine weitere Minimierung von Aufwand und Kosten kann durch Generierung von automatischen Testabläufen erzielt werden. Auf Grund dieser Forderung wurde der hier vorgestellte Evaluierungsprozess in ein bestehendes Testmanagementsystem integriert.



Testdurchführung

- Abbildung der gesamten Dialogstruktur im Testmanagementsystem
- Anreizung des Systems über entsprechende Schnittstellen (diskrete Signale, CAN, MOST,...)
- Einspielung von Sprachsamples und Hintergrundgeräusch über Mikrofonschnittstelle
- Erkennung und Auswertung der Systemreaktion über entsprechende Schnittstellen
- Ablage der Ergebnisse im Testmanagementsystem

Praxistauglichkeit des sprachbedienten Systems

Die Praxistauglichkeit des sprachbedienten Systems kann auf die beschriebene Art und Weise detailliert beurteilt werden. Fragestellungen hinsichtlich einzelner Testkriterien wie z.B. "*Sind dialektale Färbungen hinreichend gut abgebildet*?" oder "*Werden auch weibliche Stimmen hinreichend gut verstanden*?" können ohne weiteren Aufwand aus den gemessenen Daten beantwortet werden.

Literatur

[1] Abschlussbericht EU-Projekt "Onomastica" (1995)

[2] Klasmeyer,G., Herrenkind, M., Kreft, K.: Evaluierung der Übertragungsqualität von Freisprechern im Kfz, DAGA'06, Braunschweig (2006)