Hallo, ist jemand zu Hause?
Ermittlung der notwendigen Worterkennungsrate eines Smart-Home-Systems

Jan Krebber

[1] *Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Deutschland, Email: jan.krebber@rub.de*

Einleitung

Im Rahmen des europäischen IST-Projektes INSPIRE (INfotainment management with SPeech Interaction via REmote microphones and telefone interfaces) [3] wurde ein Sprachdialogsystem zur Steuerung von verschiedenen Hausgeräten (Fernseher, Videorecorder, Lampen, Rolladen, etc.) aufgebaut. Das System stellt eine intelligente und einheitliche Schnittstelle für den Benutzer dar. Um die Anforderungen an die Signalvorverarbeitung und Spracherkennung zu spezifizieren, wurden Wizard-of-Oz-Tests durchgeführt mit dem Ziel, die notwendige Worterkennungsrate für einen Keyword-Spotting-Spracherkenner zu ermitteln. Der Artikel stellt die verwendeten Methoden und die so gewonnen Ergebnisse dar.

Das INSPIRE Sprachdialogsystem

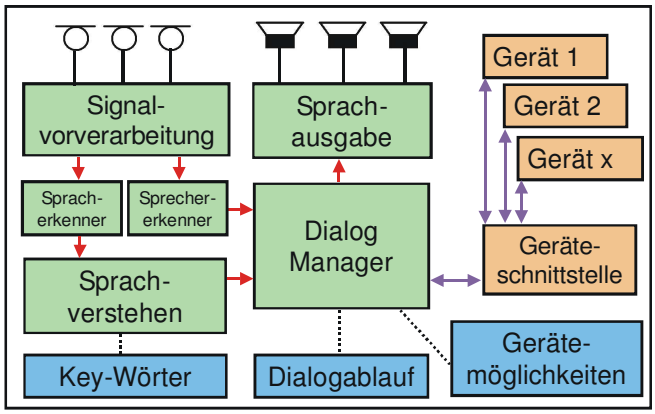

Abbildung 1: Module des INSPIRE-Sprachdialogsystems.

Das INSPIRE-Sprachdialogsystem besteht aus mehreren Modulen, die teilweise miteinander verbunden sind. Wie in Abbildung 1 zu sehen, gelangen die Sprachsignale vom Mikrofonarray zur Signalvorverarbeitungsstufe. Dort findet ein Beam-Forming und eine Störgeräuschreduktion statt. Anschließend gelangt das Signal zum einen zur Sprechererkennerstufe, zum anderen zum Spracherkenner. Der Sprechererkenner bewertet, ob der aktuelle Sprecher berechtigt ist, das System zu bedienen und gibt diese Information an den Dialogmanager. Der Spracherkenner gibt die erkannten Wörter weiter zum Sprachverstehen-Modul. Dort wird mit Hilfe von Key-Wörtern bzw. Wortgruppen versucht, eine eindeutige Zuordnung der Äußerung zu erzielen. Diese wird an den Dialogmanager weiter gegeben. Der Dialogmanager steuert den Dialog mit Hilfe der Dialogstrukturen, die in der Dialogablauf-Datenbank enthalten sind. Des weiteren benötigt der Dialogmanager die möglichen physikalischen Zustände der verwendeten Geräte; daher sind alle möglichen Zustände eines jeden Gerätes in der Gerätemöglichkeiten-Datenbank enthalten. Die verwendeten Geräte sind über eine Schnittstelle mit dem Dialogmanager verbunden. Die Schnittstelle erlaubt einen bidirektionalen Verkehr zur aktuellen Statusabfrage der einzelnen Geräte. Weiterhin steuert der Dialogmanager die für die Sprachausgabe notwendigen Lautsprecher und generiert die passenden Antwortsätze.

Die Wizard-of-Oz-Umgebung

In der Wizard-of-Oz-(WoZ)-Umgebung können die Signalvorverarbeitung, der Spracherkenner, das Sprachverstehen und Teile des Dialogmanagers ersetzt werden. Der Funktionsumfang der zu ersetzenden Einheiten kann je nach Aufgabenstellung für einzelne Versuche angepasst werden. Im konkreten Fall wurden das Mikrofonarray, die Signalvorverarbeitung und der Spracherkenner ersetzt. Um einen Keyword-Spotting-Spracherkenner mit unterschiedlichen Erkennungsraten zu simulieren, wurden die Äußerungen der Benutzer 1:1 verschriftet und an eine Wortvertauschungseinheit übergeben. Hier werden nach einstellbaren Bedingungen (u.a. die Wortfehlerrate) eine vorgegebene Anzahl der transkribierten Wörter mit phonetisch ähnlichen Wörtern vertauscht oder ausgelassen. Die relativen Vertauschungswahrscheinlichkeiten wurden in vorhergehenden Tests ermittelt. Um eine genaue Wortfehlerrate zu erzielen, ist es notwendig, dass der Wizard die Äußerungen der Versuchsteilnehmer fehlerfrei verschriftet. Wie in Abbildung 2 zu sehen, gelangt die neu entstandene Wortkette nun zum Sprachverstehen-Modul. Von hier an arbeitet das System wieder normal – mit einer Ausnahme: Die Sprechererkennung wurde aus Gründen einer vereinfachten Versuchsumgebung nicht verwendet.

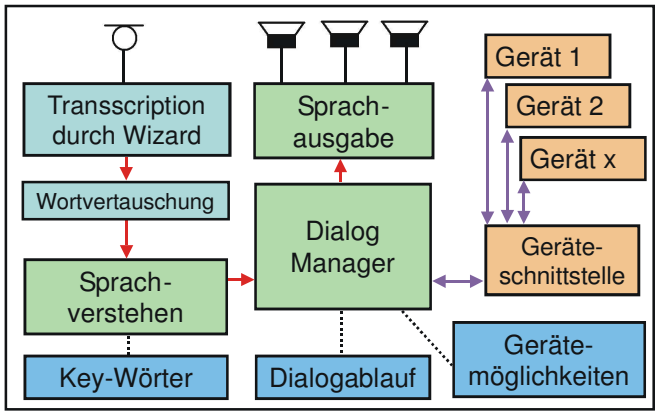

Abbildung 2: WoZ-Umgebung zur Ermittlung der notwendigen Worterkennungsrate.

Tests zur Ermittlung der minimalen Worterkennungsrate

Um den Entwicklern der Signalvorverarbeitungs- und Spracherkennermodule innerhalb des INSPIRE-Projektes eine realistische Anforderung seitens des Dialogsystems geben zu können, wurde ein Test zur Ermittlung der minimalen Worterkennungsrate durchgeführt. Für diesen Test wurde die oben beschriebene WoZ-Umgebung verwendet.

An dem Test nahmen 28 Personen im Alter von 19 –50 Jahren teil. Der Altersschnitt lag bei 26,4 Jahren. Es wurden vier unterschiedliche Ziel-Worterkennungsraten verwendet, 100%, 86%, 73%, 60%. Pro Teilnehmer wurden 3 unterschiedliche Worterkennungsraten (Szenario) getestet. Je Szenario hatten die Versuchsteilnehmer 12-14 Aufgaben zu erfüllen. Nach jedem Szenario mussten die Versuchspersonen einen Fragebogen mit 37 Fragen zu Qualitätsaspekten beantworten. Die dort verwendeten Skalen entsprechen der ITU-T Rec. P.851 [4].

Zur Bestimmung der minimalen akzeptablen Wortfehlerrate wurden nur, im mathematischen Sinne, signifikante Ergebnisse verwendet, die bei steigender Erkennungsrate einen Übergang von einer negativen (< 0) zu einer positiven Bewertung (> 0), d.h. einen Nulldurchgang zeigen. Sich nicht signifikant ändernde Beurteilungen wurden nicht berücksichtigt.

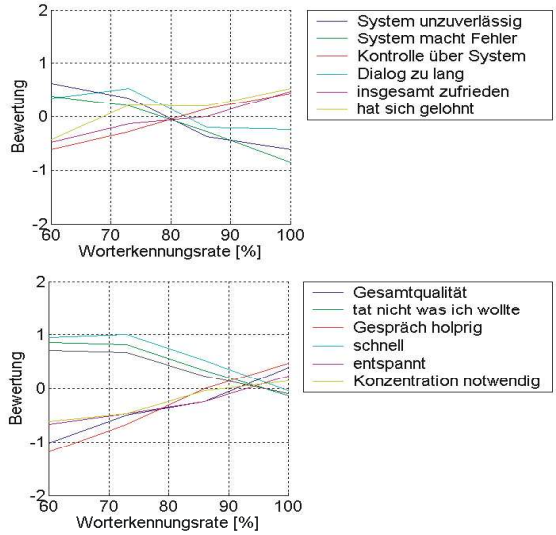

Abbildung 3: Signifikante Ergebnisse einzelner Qualitätsmerkmale, die einen Übergang von gut nach schlecht oder umgekehrt zeigen.

Wie in Abbildung 3 zu sehen, lassen sich die Ergebnisse nach ihren Übergangspunkten gruppieren. Im Bereich von 80% - 85 % liegen Qualitätsaspekte, die mit der technischen Realisierung des Systems zusammenhängen. Der Bereich zwischen 86% und 100% scheint mit allgemeinen Aspekten des Dialogs zusammenzuhängen.

Um nun einen exakteren Übergangspunkt zu finden, werden die relevanten Ergebnisse gemäß ihrer Signifikanz (signifikant, $p<0.1$, sehr signifikant, $p<0.005$, höchst signifikant, $p<0.001$) gewichtet. Mit unterschiedlichen Gewichtungen kommt man zum Ergebnis, dass die Erkennungsrate mindestens 86% - 89% betragen sollte, um mit dem INSPIRE-System eine akzeptable Qualität zu erreichen. Ein weiteres wichtiges Ergebnis dieser Tests war, dass das INSPIRE-System auch noch mit Erkennungsraten von 60% zu bedienen ist. Selbst mit der geringen Erkennungsrate waren die Teilnehmer in der Lage, die an sie gestellten Aufgaben zu lösen

Zusammenfassung

Im Rahmen des INSPIRE-Projektes wurden Tests zur Ermittelung der minimalen Wortfehlerrate durchgeführt, mit der eine noch akzeptable Qualität zu erzielen ist. Die Ergebnisse dienten zur späteren Optimierung des Spracherkenner-Moduls und des Signalvorverarbeitung-Moduls.

Die Ergebnisse zeigen eine unterschiedliche Gewichtung in Bezug auf ihren Ursprung. Positive Bewertungen zur technischen Realisierung des Systems erfordern eine geringere Worterkennungsrate (80%-85%), während dialogbezogene Bewertungen eine höhere Worterkennungsrate (85%-100%) verlangen. Für das INSPRE-System ist eine minimale Worterkennungsrate von 86% - 89% notwendig, um eine akzeptable Qualität zu erreichen. Das System ist jedoch so konzipiert, dass es auch mit Worterkennungsraten von 60% bedient werden kann.

Danksagung

Ein besonderer Dank geht an Dr. Paula Smeele von TNO, die am 08.03.2005 nach schwerer Krankheit verstorben ist. Sie war auch nach Beendigung des Projektes ein kooperativer und produktiver Partner. Die Arbeit wurde am Institut für Kommunikationsakustik (PD U. Jekosch, PD S. Möller, Prof. R. Martin) an der Ruhr-Universität Bochum angefertigt und mit Mitteln des EU-Projektes ISPIRE (IST-2001-32746) unterstützt. Vielen Dank auch an Rosa Pegam und Anders Krosch, die die Durchführung der Experimente tatkräftig unterstützt haben.

Literatur

[1] Bernsen, N.O., Dybkjær, H., Dybkjær, L. (1998). *Designing Interactive Speech Systems: From First Ideas to User Testing*, Springer, D-Berlin.

[2] Fraser, N.M., Gilbert, G.N. (1991). *Simulating Speech Systems*, Computer Speech and Language **5**, 81-99.

[3] INSPIRE. www.inspire-project.org .

[4] ITU-T Recommendation P.851 (2003). *Subjective Quality Evaluation of Telephone Services Based on Spoken Dialogue Systems*, International Telecommunication Union, CH-Geneva.

[5] Möller, S. (2005). *Quality of Telephone-Based Spoken Dialogue Systems*, Springer, US-Boston MA.

[6] Smeele, P., Möller, S., Krebber, J., Pegam, R., El Mehelmi, N., Boland, H., Hoonhout, J., Schuchardt, D., Melichar, M. (2004). *System Acceptability Evaluation Report,* Deliverable 6.3, IST project INSPIRE (IST-2001-32746), TNO Human Factors, The Netherlands.