

Projektgruppe für Hör-, Sprach- und Audiotechnologie des Fraunhofer IDMT

Die im August 2008 in Oldenburg gegründete Projektgruppe für Hör-, Sprach- und Audiotechnologie des Fraunhofer IDMT verbindet die weltweit anerkannte Oldenburger Hörforschung mit den am Fraunhofer IDMT entwickelten Kompetenzen und Technologien im Bereich digitaler Medien. Als Partner im Exzellenzcluster »Hearing4all« ist es Ziel der Projektgruppe, die wissenschaftlichen Erkenntnisse aus der universitären Grundlagenforschung in neue Technologien umzusetzen. Mit ihren Forschungsschwerpunkten adressieren die Wissenschaftler Kunden in den Branchen Telekommunikation, Multimedia, Gesundheit, Verkehr und Sicherheitstechnologie.

Fraunhofer-Institut für Digitale Medientechnologie IDMT

Projektgruppe
Hör-, Sprach- und Audiotechnologie
Haus des Hörens
Marie-Curie-Straße 2
26129 Oldenburg

Telefon +49 441 2172-400
Fax +49 441 2172-450

Ansprechpartner

Stefan Goetze
s.goetze@idmt.fraunhofer.de

www.idmt.fraunhofer.de/hsa

Ambiente Sprachsteuerung



Die skalierbaren Spracherkennungssysteme der Projektgruppe für Hör-, Sprach- und Audiotechnologie werden an unterschiedliche Einsatzszenarios angepasst und können problemlos in bestehende Anwendungen und Endgeräte integriert werden.



Mithilfe von ambienter Sprachsteuerung können auch technikferne Nutzer von komplexen technischen Systemen profitieren. Besonders wichtig ist dies bei Assistenzsystemen für ältere oder körperlich eingeschränkte Menschen. Die lautbasierte Spracherkennung der Projektgruppe für Hör-, Sprach- und Audiotechnologie kann sogar an die verminderte Artikulation von Schlaganfallpatienten angepasst werden. Ein weiteres Einsatzgebiet von Spracheingabesystemen im Bereich der Pflege ist die Unterstützung des Personals bei der Dokumentation.

Bild: Thinkstock

Sprachsteuerung

Intelligente Spracheingabesysteme kommen zunehmend als Bedienschnittstelle in der Mensch-Technik-Interaktion zum Einsatz. Die Projektgruppe für Hör-, Sprach- und Audiotechnologie arbeitet an Spracherkennungssystemen, die selbst bei Umgebungslärm und bei einem größeren Abstand zwischen Mikrofon und Sprecher robust funktionieren. Einsatzgebiete sind die komfortable Bedienung von Mediensystemen und Gebäudetechnik, die sichere Steuerung von Funktionen im Fahrzeug oder die berührungslose und zuverlässige Bedienung von technischen Geräten in rauer oder steriler Umgebung.

Robuste Erkennung im Fernfeld

Die in der Projektgruppe entwickelten Verfahren zur Erkennung von Sprachbefehlen zeichnen sich durch eine hohe Robustheit gegenüber Störgeräuschen und Raumeinflüssen aus. Die Signalverarbeitung im auditorischen System des Menschen ist Vorbild bei der Entwicklung von Erkennersystemen, die Sprache und andere Laute differenziert unterscheiden können. Aktuelle wissenschaftliche Erkenntnisse aus der psychoakustischen und psychophysikalischen Grundlagenforschung werden herangezogen, um Algorithmen mit minimaler Fehl-Erkennungsrate zu entwickeln. Zusätzlich optimiert wird die Erkennerleistung, durch Berücksichtigung der akustischen Umgebungsbedingungen. Bei ambient im Raum verbauten Mikrofonen werden Sprachsignale so auch bei Umgebungsgeräuschen wie Verkehr, Rasenmähen, laufenden Mediengeräten oder Raumhall zuverlässig erkannt.

Skalierbares Vokabular

Das verwendete Vokabular zur Sprachsteuerung ist frei definierbar und kann an individuelle Anwendungen angepasst werden – von wenigen Befehlen für die Steuerung einfacher technischer Systeme bis zu dem Wortschatz eines verstehenden, dialogorientierten Robotersystems. Das Vokabular beinhaltet auch ein individuelles Wortfolgemodell, in dem die Wahrscheinlichkeit definiert ist, mit der ein Wort auf ein anderes folgt. Durch Anpassung des Vokabulars an die Anforderungen des Einsatzgebietes wird eine hohe Erkennungsgenauigkeit erreicht.

Optimierte Signalaufnahme

Die Projektgruppe entwickelt verschiedene Technologien zur Signalaufnahme und -verbesserung, die modular kombiniert und integriert werden können. Je nach akustischen Anforderungen kommt eine individuelle Mikrofonierung bis hin zu ambient verbauten Arrays mit einer hohen Anzahl von Mikrofonen unterschiedlichster Bauart zum Einsatz. Durch ein- oder mehrkanalige Signalvorverarbeitung wie Richtungsfilterung (Beamforming), Echoreduktion, Nachhallunterdrückung oder Signalverzerrung wird zunächst das Signal für die Verarbeitung im Spracherkennungsoptimierer optimiert. Mithilfe von FPGA-basierten Hardwarekomponenten ist eine nahezu latenzfreie Übertragung per Internet Protokoll auch für mehrkanalige Signale und über weite Strecken möglich, z. B. für die isosynchrone Übertragung zur Lokalisation in Sensornetzwerken.

Signalverarbeitung auf eingebetteten Systemen

Die Spracherkennungsalgorithmen des Fraunhofer IDMT sind auf unterschiedlichen Plattformen lauffähig, so dass eine Einbindung in bestehende Applikationen und Endgeräte problemlos möglich ist. Bei Umsetzung der Erkennersysteme auf DSP oder ARM-Prozessoren kann die gesamte Verarbeitung im Gerät erfolgen, eine Verbindung zum Internet ist nicht erforderlich. Dadurch ist ein Einsatz auch in Bereichen mit hohen Anforderungen an Datenschutz und Privatsphäre möglich.

Auftragsforschung

- Entwicklung von anwendungsspezifischen Erkennertechnologien
- Anpassung des Vokabulars
- Mikrofonierung und Anpassung an akustische Einflüsse
- Verfahren der Signalvorverarbeitung
- Implementierung auf eingebettete Systeme
- Technische Evaluation von am Markt erhältlichen Systemen



Von Komfort bis Sicherheit – ambiente Sprachsteuerung ermöglicht die blickfreie und berührungslose Bedienung von technischen Geräten. Einsatzszenarien reichen von der Steuerung von Funktionen im Auto, von Haushalts- und Mediengeräten und Gebäudetechnik im Smart Home bis zu mobilen Applikationen (»Apps«) und Notrufsystemen. Bild: Thinkstock