

© 3ffi, shutterstock.com

Zum Diktat bitte!

Neural Speech: Informationstechniker der HTW Dresden entwickeln gemeinsam mit der Linguwerk GmbH eine Software zur Spracherkennung, die Nutzer an ihr Fachgebiet anpassen können.

Birte Urban-Eicheler

Schnelle Rechner und funktionierende Technik machen es möglich: Spracherkennung am PC und an mobilen Endgeräten ist längst in unserem Alltag angekommen. Sie hilft, Aufgaben schneller und komfortabler zu erledigen, vor allem, wenn gesprochenes Wort automatisch in Echtzeit in ein Schriftdokument umgewandelt wird. Denn die meisten Menschen sprechen etwa drei Mal schneller, als sie mit Maus und Tastatur arbeiten. In den Markt für Spracherkennungssysteme, der derzeit von wenigen amerikanischen Hightech-Unternehmen dominiert wird, wollen Professorin Gudrun Flach und ihr Team im Verbund mit der Dresdner Linguwerk GmbH vorstoßen. Seit Mai dieses Jahres erforschen und entwickeln sie gemeinsam eine von den Nutzern selbst anzupassende Software. Der Freistaat Sachsen fördert das Projekt für knapp drei Jahre aus Mitteln des Europäischen Fonds für regionale Entwicklung (EFRE).

Das Vorhaben des mittelständischen Technologieunternehmens und des Teams der Professorin für Grundlagen der Elektrotechnik/Technische Informatik hört sich erst einmal sehr ambitioniert an. Aber die Partner machen es anders als

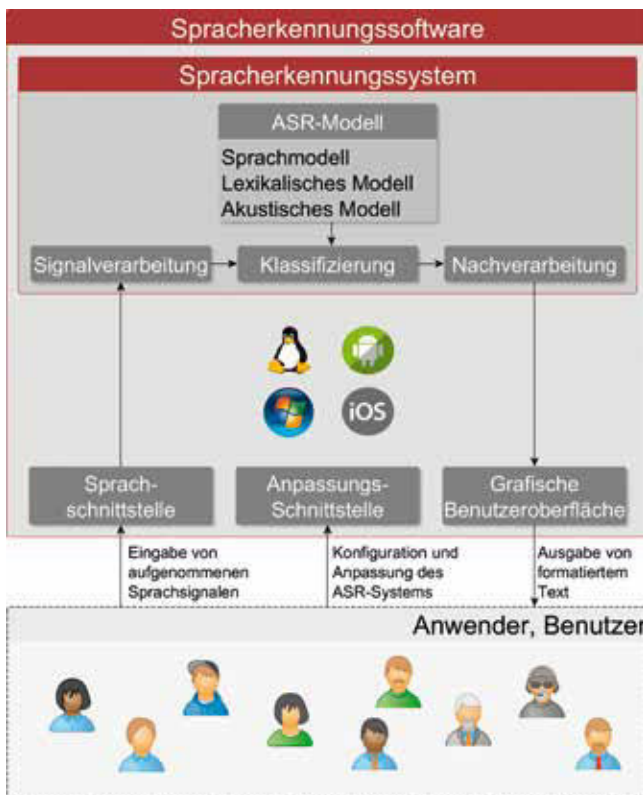
die großen Marktführer, denn Spracherkennung funktioniert mittlerweile bei Alltagssprachlichen Texten ohne Fachvokabular und komplizierte Eigennamen ziemlich gut. „Wir wollen den Kunden ein Standardsystem anbieten, das sie ohne technische Vorkenntnisse ganz selbstständig mit dem speziellen Vokabular für ihr Fachgebiet anpassen können, egal ob es sich dabei um juristische, medizinische oder technische Begriffe handelt“, erläutert Jörg Knorr von Linguwerk. „Für eine individuell zugeschnittene Adaption der Spracherkennung müssen sie kein Unternehmen mehr extra beauftragen.“

Daten bleiben vor Ort

Interessant könnte die Dresdner Spracherkennungssoftware auch für Kunden sein, die mit sensiblen Daten arbeiten, beispielsweise Behörden, die Polizei, Gerichte, Rechtsanwälte oder Mediziner, und die derzeit aus datenschutzrechtlichen oder sicherheitsrelevanten Gründen keine Spracherkennung einsetzen. „Aktuell arbeiten Diktiersysteme mit Spracherkennung meist cloudbasiert“, so Gudrun Flach. „Die Nutzerdaten gehen zu einem entfernten Server, dessen Standort in der

Regel unbekannt ist, werden dort verarbeitet und als Ergebnis kommt ein Text in Schriftform an den Nutzer zurück.“ Die Dresdner Software hingegen setzt auf die lokale Verarbeitung direkt auf einem vom Kunden selbst betriebenen Rechner oder Server. Der Kunde soll zudem in die Lage versetzt werden, sie selbstständig mit neuen oder ergänzenden Fachbegriffen zu erweitern sowie an spezielle Bedingungen wie den Sprecher oder Geräusche in der Umgebung anzupassen.

Kernstück der neuen Dresdner Software ist eine eigens entwickelte automatische Spracherkennung, kurz ASR für Automatic Speech Recognition. Ein akustisches Modell wird hierfür mit einem Sprachmodell verbunden. Das akustische Modell erstellt aus dem Gehörten eine Symbolfolge von Lauten. Was diese Laute aussagen, welche Botschaften sie vermitteln, findet das Sprachmodell heraus. Die Modellierung stützt sich dabei auf die modernste Technologie aus dem Bereich des maschinellen Lernens. Neuronale Netze stehen im Fokus für die zu entwickelnden Algorithmen und Lernverfahren. Ein weiterer Schwerpunkt liegt in der Entwicklung von Technologien, die dem späteren Nutzer die automatisierte Anpassung auf sein Anwendungsszenario ermöglichen.



Mehrwert durch Anpassungsmöglichkeit

Das Vorhaben ist eine große Herausforderung. Schließlich sollen die Nutzer als großen Mehrwert eine Technologie erhalten, die sie ohne Unterstützung selbst weiterentwickeln und an ihre Bedürfnisse anpassen können. Dafür gilt es, eine einfach zu bedienende grafische Benutzeroberfläche zu entwickeln, die diese Anpassung intuitiv ermöglicht. Das ist das Ziel und gleichzeitig der Wettbewerbsvorteil der beiden Verbundpartner. Marktführer auf diesem Gebiet bieten Spracherkennungssysteme lediglich für ausgewählte Anwendungsbereiche an. Die im Projekt zu erforschende und zu entwickelnde Spracherkennungstechnologie kann letztlich aber in allen Bereichen eingesetzt werden, das heißt auch in denen, die von den Marktführern nicht berücksichtigt werden.

Professorin Gudrun Flach und die Linguwerk GmbH, die für zwei Jahre bis Juni 2012 in der Gründungsschmiede der HTW Dresden ansässig war, arbeiten nicht das erste Mal zusammen. Gemeinsam mit dem Technologieunternehmen entwickelten die Wissenschaftlerin und ihr Team ein Hightech-Spielzeug, das die frühkindliche Sprachentwicklung fördert. Dabei steckt in einer Plüschfigur als Schlüsseltechnologie eine automatisierte Spracherkennung, die komplett offline auf einem Mikrochip arbeitet. Das 2017 auf den Markt gekommene Spielzeug ist unter www.lingufino.de erhältlich und beherrscht mittlerweile auch Fremdsprachen.

Wie beim Vorgängerprojekt setzen die Verbundpartner auf kurze Wege, da die Technologieentwicklung sehr komplex und eng verzahnt ist. Die drei Wissenschaftler aus dem Team um Gudrun Flach arbeiten am Firmenstandort Tür an Tür mit den Industrieexperten zusammen. Ein Gewinn für beide Seiten, denn am Ende steht nicht nur eine komplexe Software als Produkt. Das Projekt ermöglicht es einem größeren Team in Sachsen erstmalig, sich auf internationalem Spitzenniveau mit automatischer Spracherkennung zu etablieren und den Stand der Wissenschaft im Bereich der ASR-Technologien sowie die automatisierte Anpassung an die Bedürfnisse von Kunden voranzutreiben.

Kontakt

Fakultät Elektrotechnik
Prof. Dr. Gudrun Flach
gudrun.flach@htw-dresden.de

Linguwerk GmbH
Dr. Rico Petrick
rico.petrick@linguwerk.de

Das zu entwickelnde Gesamtsystem in Neural Speech (Spracherkennungssystem mit Nutzerinterface): Eine in das Nutzerinterface integrierte Anpassungsschnittstelle erlaubt es, das System zur automatischen Spracherkennung (ASR) ohne Expertenwissen selbstständig zu konfigurieren und individuell anzupassen. Das ASR-System soll dabei plattformunabhängig sein, um die Nutzung auf allen gängigen Computersystemen (Desktop und mobil) zu ermöglichen.