

Doppelgänger für das Gehör

Sprachassistenten verschiedener Anbieter stehen in vielen Haushalten und sind auch auf dem Smartphone ständig dabei. Die Künstliche Intelligenz (KI), auf der sie basieren, nutzen der Kommunikationsakustiker und Exzellenzcluster-Forscher Bernd T. Meyer und sein Team für ihre Hörforschung. Dabei machen sie sich auch die vermeintlichen Schwächen dieser Sprachassistenten zunutze – denn vielfach haben sie ähnliche Schwierigkeiten wie Menschen.

Von Sonja Niemann



Das Hackfleisch für die Bolognese brutzelt in der Pfanne, das Radio dudelt und die Dunstabzugshaube heult, als die Spaghetti aus der knisternden Verpackung in den Topf platschen. Jetzt nur noch sicherstellen, den perfekten „al-dente-Zeitpunkt“ nicht zu verpassen.

„Computer, stelle einen Timer auf neun Minuten.“
 „Das weiß ich leider nicht.“
 „Computer! Timer auf neun Minuten!“
 „Es ist kein Timer gestellt.“

Es gibt Situationen, da scheinen die heimischen Sprachassistenten Befehle einfach nicht verstehen zu wollen. Trotz oder gerade wegen dieser

Schwächen haben die kleinen Geräte und die KI, auf der sie beruhen, einen besonderen Wert für die Hörforschung. Denn beim Verstehen von Sprache bei lauten Umgebungsgeräuschen haben sie ähnliche Probleme wie Schwerhörige. Daraus einen Nutzen zu ziehen, ist aber nur ein Ansatz, den Forschende des Oldenburger Exzellenzclusters Hearing4all verfolgen, um Spracherkennungssoftware für besseres Hören nutzbar zu machen.

Maßgeblich daran beteiligt ist Prof. Dr. Bernd T. Meyer mit seinem Team der Abteilung Kommunikationsakustik. Gemeinsam mit fünf Doktorandinnen und Doktoranden forscht er an der Schnittstelle zwischen Sprach- und Hörforschung. Seit seinem Physikstudium ist er von den Möglichkeiten moderner Spracherkennung fasziniert.

Damals schlug ihm Prof. Dr. Dr. Birger Kollmeier, heute Sprecher des Exzellenzclusters, das Thema für einen Vortrag vor. „Mich hat interessiert, warum die Software so viel schlechter Sprache erkennt als Menschen mit gesundem Gehör“, sagt Meyer. Er begann, Methoden zu entwickeln, die Systeme besser zu machen – und fand Möglichkeiten, diese Erkenntnisse in die Hörforschung zu überführen.

Ein Forschungsansatz liegt darin, Spracherkennungssoftware dafür einzusetzen, Höreinschränkungen zu diagnostizieren. Eine erste Einschätzung über das eigene Hörvermögen können Interessierte dank der Forschung aus Oldenburg sogar im eigenen Wohnzimmer erhalten, wenn sie das Smart-Home-System Alexa nutzen. Wer „Alexa, starte Hörtest“ sagt, startet den

Das menschliche Gehör nimmt akustische Signale auf, doch erst das Gehirn dekodiert Töne, Geräusche und Sprache. Sprachassistenten zerlegen Schallwellen ebenfalls in kleinere Einheiten und setzen diese wieder zu Wörtern zusammen. Dabei haben sie manchmal ähnliche Schwierigkeiten wie Schwerhörige.

sogenannten Skill aus Oldenburg, bekommt kurze Sätze vorgespielt und soll sie wiederholen.

500.000 Äußerungen haben der KI beigebracht, Sprache zu verstehen

Die Antworten erreichen den Sprachassistenten als Schallwellen und werden von ihm in kurze akustische Einheiten unterteilt. „Die KI hat zuvor gelernt, den Einheiten des akustischen Signals die kleinsten Einheiten der menschlichen Sprache, die sogenannten Phoneme, zuzuordnen“, erklärt Meyer. Damit ist das Puzzle der Spracherkennung aber nur halb gelöst:

In einem weiteren Schritt berechnet die KI die wahrscheinlichste Abfolge der erkannten Phoneme. Abhängig davon setzt sie diese im besten Fall zu dem Wort zusammen, das die Testperson gesagt hat.

Während der Hörtest für zu Hause dafür die Alexa-eigene KI nutzt, forscht das Team um Meyer normalerweise mit selbst entwickelten KIs. Die Forschenden nutzen dafür die Methode des maschinellen Lernens, bringen also einem Computer bei, in allgemeinen Daten Muster zu erkennen und so zu lernen, gesprochene Sprache in geschriebene Wörter zu übersetzen. Dafür trainieren sie ihre KI mit Sprachproben – zuletzt mit mehr als 500.000 Äußerungen von mehr als 1.000 Personen. Ein aus vielen verknüpften mathematischen Formeln bestehendes

neuronales Netz lernt anhand dieser Daten, aus einem Input (hier: einer Schallwelle) einen Output (hier: ein geschriebenes Wort) zu generieren.

Anders als der Alexa-Hörtest, der lediglich mit einem Ampelsystem Rückmeldung über eine mögliche Hörschwäche liefert, soll ihre eigene KI so genau sein, dass sie eine klinische Diagnose liefert. Diesem Ziel ist das Team schon sehr nahegekommen, der zugrunde liegende Algorithmus ist weit entwickelt. „Wir haben gezeigt, dass unser neuronales Netz mit ähnlicher Genauigkeit wie ein von einer medizinischen Fachkraft begleiteter Test zeigen kann, wie gut eine Person hört“, sagt Meyer.

Trotz stetiger Verbesserungen in Diagnose und Versorgung mit Hörgeräten ist es bis heute nicht möglich, bei

beeinträchtigten Menschen das Hören in all seinen Facetten wiederherzustellen. Denn es genügt nicht, die Geräuschkulisse zu verstärken – im Alltag geht es vor allen Dingen um das Verstehen von Sprache. Und dabei haben Menschen ganz ähnliche Probleme wie der Sprachassistent in der Küche: Zwischen den vielen Umgebungsgereuschen geht das gesprochene Wort schnell unter.

Diese gemeinsame Schwäche versucht Meyers Forschungsteam mit einem weiteren Forschungsansatz in einen Vorteil zu verwandeln. Statt wie beim Hörtest ein neuronales Netz so zu programmieren, dass es einen Sprecher besonders gut versteht, forscht der Physiker gemeinsam mit seiner Doktorandin Jana Roßbach daran, einen Sprachassistenten exakt genauso gut oder schlecht wie eine real existierende Person hören zu lassen. Große Beachtung fand kürzlich eine Studie, bei der es den Oldenburgern gelungen war, das neuronale Netz so auf die jeweilige Höreinschränkung von Probandinnen und Probanden zu kalibrieren, dass Mensch und Computer bei Hörtests nahezu identisch abschnitten – wie akustische Doppelgänger. „Die Idee dahinter ist: Wenn eine KI voraussagen kann, dass ein Hörgeräteträger ein in seiner Umgebung gesprochenes Wort nicht versteht, dann kann sie die Einstellungen des Hörgeräts auch so optimieren, dass die Person dem Gespräch doch folgen kann. Aus einer

Not machen wir also eine Tugend“, erklärt Meyer.

„Bisher haben wir ausschließlich das Hören mit nur einem Ohr untersucht. Wir wissen aber, dass gerade das binaurale Hören positive Effekte auf das Sprachverstehen hat“, ergänzt er. Deshalb soll die KI jetzt lernen, auch dann die gleichen Prognosen zu treffen wie ein Mensch, wenn dieser im Test beide Ohren nutzt.

Einsetzbar nicht nur bei Hörgeräten, sondern bei allen Hearables

Bei den bisherigen Versuchen wurde die KI mit Sprachproben gefüttert, die von einem starken Rauschen überlagert waren, das die Höreinschränkung der Testperson simulierte. Diese Sprachprobe transkribierte die KI und glich das Ergebnis mit der Information ab, was der Sprecher tatsächlich gesagt hatte. Waren beide Informationen identisch, wusste die KI, dass auch die Testperson das Sprachbeispiel verstehen würde. „Im echten Leben weiß eine KI aber natürlich nicht, was tatsächlich gesagt wird“, sagt Meyer. Gemeinsam mit Roßbach arbeitet er deshalb daran, dass eine Prognose auch ohne diese Information möglich ist. Als Grundlage dient die Tonqualität der Sprachprobe.

Der Nutzen von KI-Anwendungen wie dieser beschränkt sich nicht nur auf Hörgeräte. „Sie sind relevant für die gesamte Kommunikation mit Hearables“, sagt Meyer. Dazu zählen neben Hörgeräten auch smarte Kopfhörer, die längst mehr als nur Musik abspielen: Sie blenden Umgebungsgereusche aus oder verstärken Sprache in der Umgebung gezielt. Eine Individualisierung an die Hörfähigkeiten oder auch -vorlieben ihres Trägers könnte der nächste logische Entwicklungsschritt sein. Noch stellen der benötigte Strombedarf für komplizierte Algorithmen und die begrenzte Leistungsfähigkeit der verbauten Prozessoren Hürden dar.

Ein Beispiel aus Meyers Abteilung zeigt aber, dass solche Hürden in der Oldenburger Hörforschung bereits regelmäßig überwunden werden. 2017 bei einem wissenschaftlichen Vortrag in Baltimore war Meyer beeindruckt von einer Demonstration, bei der eine KI die Signale von zwei gleichzeitig in ein Mikrofon sprechenden Personen voneinander trennen konnte. „Das hat mich damals umgehauen“, sagt er. Obwohl dieses Beispiel ein vielversprechender Ansatz für ein Grundproblem der Hörforschung – die Trennung von erwünschten und unerwünschten Geräuschen – zu sein schien, war er gerade dafür zunächst unbrauchbar. „Die Anwendung brauchte riesige Ressourcen und zu viel Zeit. Hörgeräte dürfen mit einer maximalen Verzögerung von zehn Millisekunden arbeiten, sonst führt die Überlagerung der normalen Geräuschkulisse mit den verzögerten Hörsignalen zu Verzerrungen“, erklärt Meyer.

Mit Doktorand Nils Westhausen hat er jetzt eine Möglichkeit gefunden, den Prozess auf zwei Millisekunden zu beschleunigen – schnell und ressourcenarm genug, um perspektivisch in Hörgeräten zu funktionieren. Die Technologie könnte es Menschen ermöglichen, aktiv das akustische Signal auszuwählen, dem sie zuhören möchten – und zum Beispiel heulende Dunstabzugshauben und brutzelnde Pfannen auszublenden. Das wäre ein weiterer echter Durchbruch.

EINBLICKE 2023/24

Wie verringern wir soziale Ungleichheiten?

Ausblicke



Prof. Dr. Gundula Zoch

Soziologie sozialer Ungleichheiten

„Wir stehen vor enormen sozialen Herausforderungen, etwa durch den Wandel des Arbeitsmarktes und neuartige Technologien. Um diesen gewachsen zu sein, ist ein umfassender Zugang zu Bildung für Kinder und Erwachsene unerlässlich. Sie alle benötigen geeignete Kompetenzen, um Chancen zu nutzen und den gesellschaftlichen Wandel zu meistern.

Bildung ist zudem ein wichtiger gesellschaftlicher Klebstoff. Sie spielt nicht nur für materiellen Wohlstand und Wohlbefinden eine Rolle, sondern auch für die gesellschaftliche, politische und soziale Teilhabe. Wir benötigen jedoch dringend Reformen, um mehr Chancengerechtigkeit im Bildungssystem zu gewährleisten.

Die Forschung zeigt: Gute frühkindliche Bildung ist entscheidend, um Bildungsungleichheiten zu verringern. Kinder, auch aus benachteiligten Gruppen, profitieren davon in ihrem gesamten Lebensverlauf. Ältere Generationen gewinnen ebenfalls, da staatliche Leistungen von den Jüngeren getragen werden, die sich den technologischen Entwicklungen auf dem Arbeitsmarkt stellen und diese durch ihre Fähigkeiten nutzen oder ergänzen.

Meine Vision ist, dass wir in 50 Jahren in Deutschland ein Bildungssystem geschaffen haben, das alle mitnimmt, unabhängig von ihren familiären Bedingungen und ihrer Herkunft. Und dass wir Bildung als einen lebenslangen Prozess begreifen und so Menschen befähigen, auch in Zeiten großer Veränderungen erfolgreich teilzuhaben. Hierfür lohnt es sich, Geld und Expertise in die Hand zu nehmen, weil wir so in allen Bereichen soziale Ungleichheiten verringern können.“



Der Kommunikationsakustiker Bernd T. Meyer betreibt Sprach- und Hörforschung auf Basis des maschinellen Lernens. Eines seiner Projekte: Sprachassistenten für präzise Hörtests zu nutzen.