

# Sprachsignalen auf der Spur



Umgebungsgeräusche, Nachhall oder akustische Rückkopplungen erschweren oft das Verstehen von Sprache, wenn technische Geräte im Spiel sind. Der Hörforscher Simon Doclo nähert sich dem Problem mit mathematischen Verfahren – mit Erfolg

In Bruchteilen von Sekunden quer durch das Weltall kommunizieren, das machte der „Kommunikator“ in der legendären Star-Trek-Serie schon Ende der 1960er-Jahre möglich. Und auch wenn interstellare Kommunikation noch Zukunftsmusik ist, so verlassen wir uns längst auf technische Unterstützung – Handys, Telefonkonferenzen oder Hörhilfen – um mit unseren Mitmenschen im Gespräch zu bleiben.

Trotz rasanten Fortschritts hat die Technik, die uns bei der Kommunikation unterstützt, immer noch ihre Defizite. Alle, die etwa versuchen, in einer lärmenden Bahnhofshalle zu telefonieren oder in einem vollen Restaurant einem Gesprächspartner am anderen Ende des Tisches zu lauschen, kennen das Problem: Umgebungsgeräusche oder Nachhall machen es Normalhörenden und erst recht Menschen, die auf Hörhilfen angewiesen sind, schwer, das Gesprochene zu verstehen. Experten sprechen auch vom sogenannten Cocktail-Party-Effekt – ein Begriff, der bereits in den 1950er-Jahren geprägt wurde.

Hier setzt die Forschung von Prof. Dr. Simon Doclo an, der am Department für Medizinische Physik und Akustik der Universität die Abteilung „Signalverarbeitung“ leitet. „Wir wollen die Sprachkommunikation in schwierigen akustischen Umgebungen verbessern – und zwar dann, wenn wir Geräte wie Hörhilfen oder Handys nutzen“, sagt er. Sein Team stellt sich der Herausforderung, der sich die Oldenburger Hörforscher im Exzellenzcluster „Hearing4all“ und im Sonderforschungsbereich (SFB) „Hörakustik“ aus unterschiedlichsten Perspektiven widmen, mit mathematischen Methoden: Es entwickelt Algorithmen – also Abfolgen von Rechenvorschriften – mit denen sich Effekte herausrechnen oder unterdrücken lassen, die die Sprachverständlichkeit und die Sprachqualität bei der Kommunikation stören.

Seit 2009 forscht und lehrt der gebürtige Belgier an der Universität – ein Zeitpunkt, zu dem die Oldenburger

Hörforschung bereits Weltruf hatte, wie er betont. In seiner Forschung geht Doclo üblicherweise in drei Schritten vor: Am Anfang stehe das Konzipieren eines neuen Algorithmus. „Im nächsten Schritt implementieren und optimieren wir den neuen Algorithmus und untersuchen mithilfe von Computersimulationen, ob dieser den gewünschten Effekt erzielt, also beispielsweise die Sprachqualität um einen bestimmten Prozentsatz verbessert“, erläutert der Elektroingenieur.

Um nicht nur ein technisches Maß zu haben, testen die Wissenschaftler im letzten Schritt in Hörversuchen mit Probanden, ob das neue Vorgehen tatsächlich funktioniert – eine aufwendige Aufgabe, die oft gemeinsam mit anderen Gruppen im SFB und Exzellenzcluster durchgeführt werde, betont Doclo. Was nach einer routinemäßigen Aufgabe klingt, ist hochkompliziert: Denn die Arbeit mit mathematischen Verfahren, die helfen sollen, aus komplexen Schallinformationen gut verständliche Sprache herauszufiltern, ist oft sehr mühsam. „Nur ein Teil der Algorithmen, die wir entwickeln, ist letztlich so gut, dass sie für die abschließenden Tests mit Probanden geeignet sind“, sagt Doclo.

### „Der ‚heilige Gral‘ der akustischen Signalverarbeitung.“

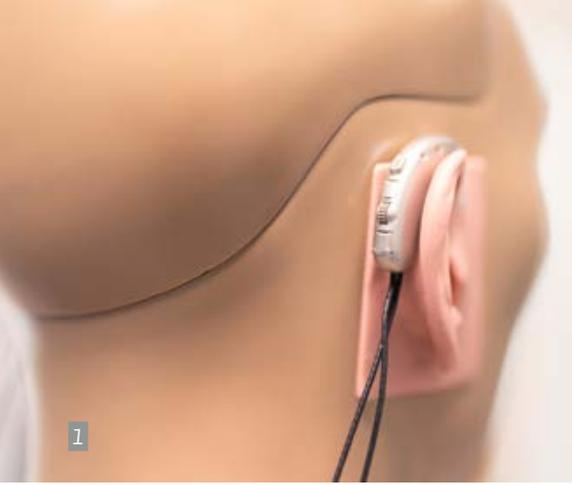
In seiner Arbeit macht sich der 46-jährige zunutze, dass die meisten Geräte, wie etwa Hörhilfen, mehrere Mikrofone enthalten. Das heißt, die Forscher können aus einem Schallsignal nicht nur die sogenannten spektralen Komponenten herausfiltern – also etwa spezielle Frequenzen mit mehr oder weniger Hintergrundgeräuschen. „Die verschiedenen Mikrofonensignale geben außerdem Aufschluss darüber, wie sich der Schall in einem Raum ausbreitet“, erläutert Doclo.

Um, wie er es selbst formuliert, kluge Algorithmen zu entwickeln, setzt Doclo gemeinsam mit seinem Team auf zwei unterschiedliche Herange-

hensweisen: „Einerseits arbeiten wir mit klassischen Methoden der digitalen Signalverarbeitung und nutzen statistische Modelle von Sprache und den akustischen Eigenschaften eines Raums aus“, sagt er. Das heißt, dass die Forscher etwa die Frequenzen eines akustischen Signals, also die Schwingungen des Schalls, und ihre Änderungen über die Zeit mit statistischen Methoden beschreiben.

Mithilfe solcher Zeitfrequenzanalysen beispielsweise werden die optimalen Parameter einer mathematischen Zielfunktion geschätzt, um ein sauberes Sprachsignal aus verrauschten und verhallten Aufnahmen herauszufiltern. „Eine wesentliche Herausforderung dabei ist, eine geeignete Zielfunktion zu konzipieren, die sich einerseits relativ leicht optimieren lässt und andererseits die psycho-akustischen Eigenschaften des menschlichen Gehörs umfasst“, sagt Doclo. Zudem sollten die entwickelten Algorithmen auch blind funktionieren: Denn bei klar definierten Laborexperimenten können die Forscher zwar annehmen, dass die genaue Position einer Schallquelle oder die Mikrofonabstände bekannt sind. „Doch in der Praxis sind diese Größen oft unbekannt“, erläutert der Forscher.

Einen wichtigen Fortschritt haben die Wissenschaftler in den vergangenen fünf Jahren auf diese Weise etwa im Umgang mit Nachhall erzielt. Unter Nachhall verstehen Experten die wiederholten Reflexionen eines Schallsignals, die entstehen, wenn der Schall auf Hindernisse wie Wände trifft und mehrfach zurückgeworfen wird. „Das ist der ‚heilige Gral‘ der akustischen Signalverarbeitung, denn es ist sehr schwierig, die Reflexionen vom Schallsignal selbst zu trennen“, sagt Doclo. „Hier haben wir neue Verfahren entwickelt, die den Nachhall deutlich besser als zuvor herausfiltern können.“ Eine Arbeit, die auch in Fachkreisen Beachtung fand: Ende 2019 erhielt Doclo gemeinsam mit seiner Kollegin Dr. Ina Kodrasi den jährlichen Publikationspreis der Informationstechnischen Gesellschaft (ITG).



1



2



3



4

Neben den statistischen modellgetriebenen Methoden arbeitet das Team andererseits mit sogenannten datengetriebenen Verfahren. In diesen „machine-learning“-Ansätzen trainieren die Wissenschaftler ein sogenanntes neuronales Netz mit einer großen Menge an Daten. „Vereinfacht gesagt, füttern wir das Netzwerk mit sehr vielen Audiosignalen, die wir entweder unter kontrollierten Bedingungen im Labor aufgenommen oder am Rechner simuliert haben“, erläutert Doclo. Beispielsweise können die Forscher Sprache und Geräusche getrennt aufnehmen und so ein sauberes Sprachsignal als Ergebnis vorgeben. Das Netzwerk soll dann lernen, das entsprechende Signal aus den verrauschten Daten wieder herauszufiltern.

Doch Doclo betont, dass viele datengetriebene Methoden oft wie eine „Black Box“ seien. Um die eigentlichen Prozesse besser zu verstehen, sind seiner Ansicht nach daher eher die statistischen modellgetriebenen Methoden hilfreicher. „Das ist einer der Gründe, warum wir die datengetriebenen und statistisch modellgetriebenen Methoden kombinieren und so die Vorteile beider Ansätze nutzen können“, sagt er. Auf diese Weise hofft der Forscher eine möglichst gute Balance zwischen der Leistungsfähigkeit (performance) eines Algorithmus und dessen Robustheit (robustness) zu erhalten. Unter Robustheit verstehen Forscher die Eigenschaft des Verfahrens, sich so gut wie möglich verallgemeinern zu lassen, also auf viele verschiedene Situationen anwenden

zu lassen. „Unsere Algorithmen sollen ja auch für unbekannte Hörumgebungen und unbekannte Geräusche funktionieren“, sagt Doclo.

Eine weitere Herausforderung bei der Forschung des Teams ist, dass alle ihre Ansätze in Echtzeit funktionieren müssen. Denn natürlich sollen Geräte für die Sprachkommunikation, wie Hörgeräte oder Handys, das Signal für die Betroffenen nicht spürbar verzögern. „Wir müssen also die Signale, sobald sie das Gerät erreichen, möglichst schnell verarbeiten, das heißt das saubere Signal innerhalb von wenigen Millisekunden herausfiltern“, erläutert Doclo. Auch aus diesem Grund arbeiten die Forscher daran, dass ihre Algorithmen nicht zu komplex werden.

Das Team setzt die Algorithmen unter anderem auch für sogenannte akustische Sensornetzwerke ein – wenn etwa Informationen aus räumlich verteilten Mikrofonen, wie die eines Hörgeräts und die eines Handys, genutzt werden. Beispielsweise könnten Schwerhörende oder auch Normalhörende künftig in einem lauten Restaurant ein Handy, dessen Mikrofon mit den Hörgeräte-Mikrofonen zusammenarbeitet, einfach auf den Tisch legen. „Alle verfügbaren Mikrofone analysieren dann die Umgebung akustisch“, erläutert Doclo. Und auch wenn die Position der Geräte nicht bekannt ist, gelinge es in solchen Situationen immer besser, das gewünschte Sprachsignal herauszufiltern.

Doch die mathematischen Methoden stoßen auch an ihre Grenzen: „Bei

- 1 In ihren Experimenten verwenden Doclo und sein Team auch sogenannte Kunstköpfe. Diese bilden so gut wie möglich die akustischen Eigenschaften eines menschlichen Kopfes nach.
- 2 Mit einem Lautsprecher spielen die Forscher bestimmte Geräusche ab. Die Mikrofone in den Ohren des Kunstkopfes zeichnen diese auf.
- 3 Mithilfe von Soundkarten nehmen die Wissenschaftler unter anderem unterschiedliche Mikrofonensignale auf. Diese verwenden sie dann in weiteren Experimenten.
- 4 Sogenannte Mikrofonarrays kommen unter anderem in akustischen Sensornetzen zum Einsatz. Die Experten können mit ihnen die akustische Umgebung, etwa in einem Raum, analysieren.

vielen Anwendungen analysieren wir erst die akustische Umgebung, um alle relevanten Schallquellen und ihre Position zu bestimmen“, sagt Doclo. Diesen Schritt nennen Experten Auditorische Szenenanalyse (Computational Auditory Scene Analysis). Doch auf Grundlage der akustischen Signale allein ist es nicht möglich, zu entscheiden, welcher Schallquelle ein Nutzer eigentlich zuhören möchte. Hier sind ganz andere Informationen nötig. Doclos Team arbeitet daher mit dem Oldenburger Neuropsychologen Prof. Dr. Stefan Debener zusammen. Ein langfristiges Ziel der Forscher ist, mithilfe von EEG-Messungen herauszufinden, auf welche Schallquelle eine Person ihre Aufmerksamkeit lenkt, um diese Quelle gezielt zu verstärken.

### **„Wir wollen die Dinge besser machen.“**

Ein anderes kniffliges Problem, dem Doclo mathematisch auf der Spur ist, sind akustische Rückkopplungseffekte. Das typische Pfeifen einer Feedbackschleife ist fast jedem bekannt,

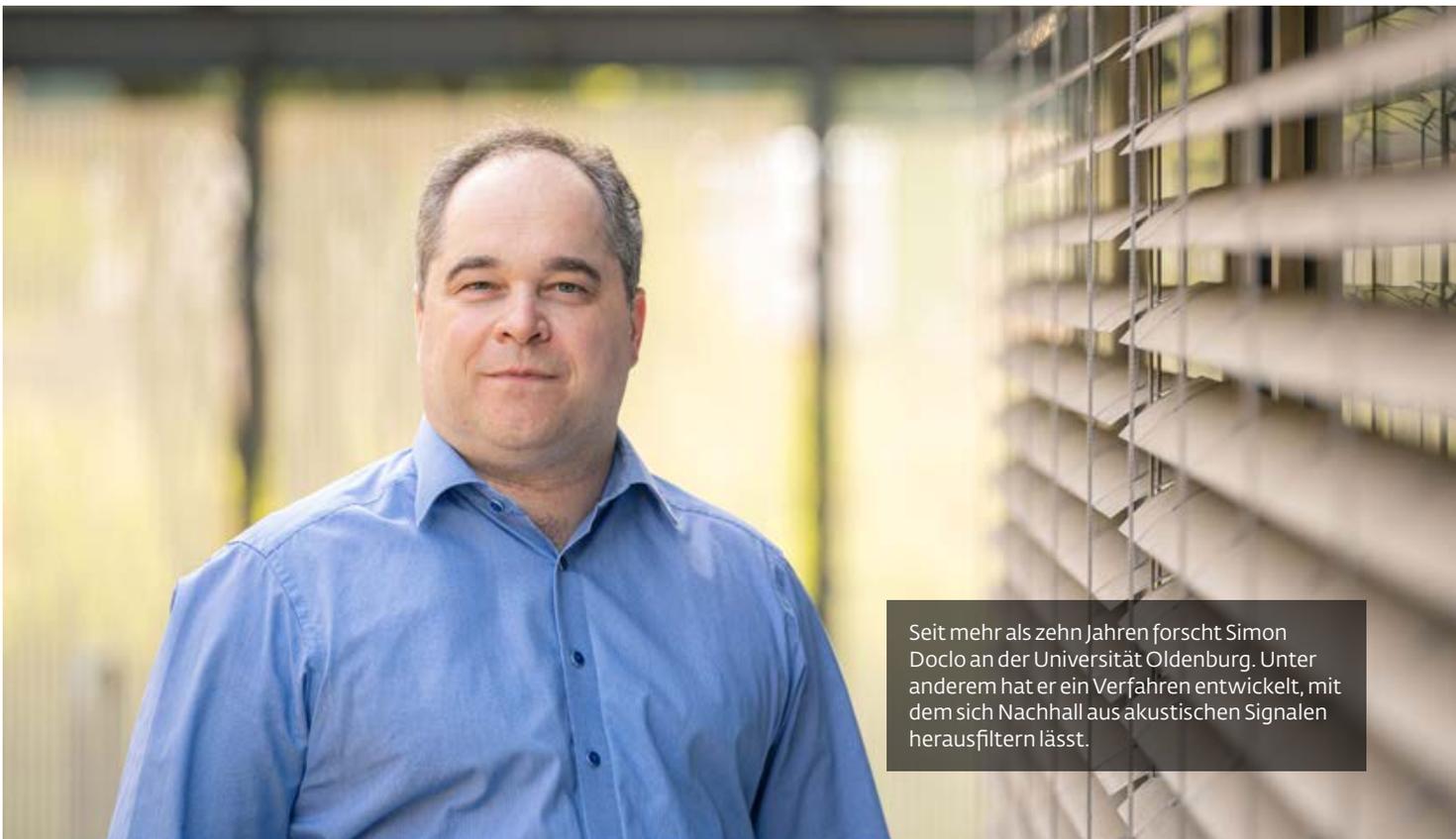
der schon einmal ein Mikrofon in die Nähe eines daran angeschlossenen Lautsprechers gebracht hat. Dieses Phänomen tritt auch in sogenannten Hearables auf. Die kleinen, im Ohr getragenen Geräte arbeiten mit Algorithmen der digitalen Signalverarbeitung und leisten weit mehr als normale Kopfhörer. Sie unterstützen Normalhörende ganz individuell darin, ein Gespräch in einer geräuschvollen Situation besser zu verstehen. Mehrere Arbeitsgruppen am Department und am Fraunhofer Institutsteil Hör-, Sprach- und Audiotechnologie (IDMT) in Oldenburg arbeiten derzeit daran, Hearables zu optimieren.

„Im Sonderforschungsbereich Hörakustik haben wir ein neues Hearable entwickelt, das zwei bis drei winzige Mikrofone im kleinen Ohrstück enthält“, erläutert Doclo. Doch da diese Mikrofone sehr nah an einem der Lautsprecher liegen, hatten die Forscher immer wieder mit Rückkopplungen zu kämpfen. „Wir haben ein vergleichsweise einfaches Verfahren entwickelt, um diese Rückkopplungen zu unterdrücken“, berichtet der Elektroingenieur. Das Verfahren habe sogar viel

besser als erwartet gearbeitet. „Das war eine Überraschung und auch ein Glücksmoment für uns.“

Daneben sieht Doclo für die Zukunft viele weitere spannende Aufgaben. Vor allem möchte er die Algorithmen durch die Kombination von statistischen modellgetriebenen und datengetriebenen Verfahren weiter optimieren. „Wir sind Ingenieure“, sagt er, „wir wollen die Dinge besser machen.“ Langfristig kann er sich vorstellen, seine mathematischen Methoden für neue Anwendungen zu erschließen – beispielsweise in der Medizin: „Auch hier gibt es verteilte Sensoren am Körper, die verrauschte Signale liefern.“ Ebenso wie in der Akustik könnten die Algorithmen helfen, gewünschte Informationen herauszufiltern.

Sein Herz schlägt aber vor allem für die Akustik, sagt Doclo – von Hörgeräten über Handys bis hin zu Smart Speakers. Sein Traum sind Verfahren, die wirklich robust sind und reibungslos funktionieren. „Die Star-Trek-Sache eben“, sagt er schmunzelnd: „Man kann herumlaufen, mit allen reden, auch über große Entfernungen, ohne dass Nebengeräusche stören.“ (cb)



Seit mehr als zehn Jahren forscht Simon Doclo an der Universität Oldenburg. Unter anderem hat er ein Verfahren entwickelt, mit dem sich Nachhall aus akustischen Signalen herausfiltern lässt.