Bild: Thorsten Hübner

Hier spricht das Hirn

KI-gestützte Sprachprothese nutzt neurale Signale

Mit Elektroden nehmen Forscher neurale Signale aus dem Gehirn auf und künstliche Intelligenz ermittelt daraus, was der Mensch sagen will. Das ist aber keine Technik für jeden. Eine Alternative bietet die Messung von Muskelreizen.

Von Arne Grävemeyer

Forscher am Cognitive Systems Lab (CSL) der Universität Bremen haben im Herbst 2021 erstmals eine echtzeitfähige Neurosprachprothese beschrieben. Die macht Wörter unmittelbar hörbar, von denen sich eine Versuchsperson lediglich vorstellt, sie würde sie sprechen. Die Neurosprachprothese schaut dabei direkt ins Gehirn, interpretiert die Hirnströme und übernimmt die Sprachausgabe. Warten Sie aber besser noch etwas, bevor Sie Ihr Mikrofon vom Headset schrauben oder im Vertrauen auf die Hirn-Computer-Schnittstelle schon Maus und Tastatur verschenken: Die Technik ist noch am Anfang und nicht für jeden zu empfehlen.

Facebooks Gedankenleser

Bereits 2017 startete Facebook in seinen Reality Labs ein Projekt, um ein „Brain-Computer Interface" (BCI) zu entwickeln. Die Entwickler hatten die faszinierende Vorstellung, dass Anwender über eine direkte Schnittstelle in aller Stille und ohne Tastatur ihre Gedanken dem Computer mitteilen können, ohne diese extra in Worte zu fassen und dann Buchstabe für Buchstabe eintippen zu müssen. Eine Arbeitshypothese war, dass der Mensch schneller denken als sprechen kann und die Texteingabe mithilfe einer solchen Schnittstelle mühelos und sogar schneller funktioniert. Man hoffte, eine Art Stirn-

band oder Kappe zu entwickeln, mit der die Träger auf Eingabe-Raten von mehr als 100 Wörtern pro Minute kommen. Kritiker befürchteten, dass die Nutzer ihre Gedanken damit dem Konzern noch unmittelbarer ausliefern würden als schon bisher durch ihre Aktivitäten auf der Social-Media-Plattform.

Der konkrete technische Ansatz der Facebook-Forscher umfasste integrierte Strahler für nahes Infrarotlicht – einen Frequenzbereich des elektromagnetischen Spektrums, der etwas langwelliger und weniger energiereich ist als das sichtbare Licht. Diese Strahler sollten Bereiche des vorderen Kortex durchleuchten. Ähnlich wie beim Scan mit einem Fingerpulsoximeter wollte man aus den durchgehenden Strahlen die Sauerstoffsättigung einzelner Hirnareale ermitteln. Aus diesen Messungen könnte man zwar niemals die gleichen Daten gewinnen wie aus Elektrodenfeldern im oder am Gehirn, die die Hirnströme ungefiltert aufnehmen, aber man hoffte doch, damit wenigstens eine Handvoll Befehlswörter sicher unterscheiden zu können.

Elektroden im geöffneten Schädel

Im Juli 2021 veröffentlichte Facebook dann einen Blogeintrag (ct.de/y779), in dem von diesen Ansätzen keine Rede mehr war. Stattdessen berichtete der Konzern, er habe im Rahmen der BCI-Forschung eine Gruppe der University of California in San Francisco unterstützt. Die hatte bereits 2019 einen Patienten vorgestellt, bei dem ein Elektrodennetz direkt auf der Großhirnrinde die Hirnströme erfasst. Deren neurale Muster nutzten die Forscher als Input für einen Machine-Learning-Ansatz. Die damit gebildete KI lernte, die Hirnströme zu interpretieren. In einem Dialogsystem gelang es den Forschern, daraus die wahrscheinlichste Antwort auf eine Frage abzuschätzen. Dem Patienten, der nicht mehr selbst sprechen konnte, gelang es durch dieses System, mit ein paar Sekunden Verzögerung auf einfache Fragen zu antworten.

In diesem Jahr konnten die Forscher in Kalifornien ihr System wesentlich verfeinern [1]. Sie implantierten inzwischen ein hochdichtes Multielektroden-Array auf dem sensomotorischen Kortex, der die Sprache steuert. Der Patient, der nach einem Schlaganfall nicht mehr selbst sprechen kann, versuchte vorgegebene Wörter zu sprechen und die dabei erzeugten neuralen Signale dienten im Deep Learning zum Training einer KI. Mit den gefundenen Mustern ist das System nun in der Lage, die gewünschten Wörter aus den Hirnströmen nahezu in Echtzeit direkt auszulesen und als Text auf einem Computerbildschirm auszugeben. Das trainierte und mittlerweile mit 98-prozentiger Sicherheit erkannte Vokabular umfasst 50 Wörter. Der Patient schaffte mit diesem System eine durchschnittliche Ausgabe von mehr als 15 Wörtern in der Minute.

Für die Forschung ist das ein schöner Erfolg, der beweist, dass sich aus den Hirnströmen eines Menschen die beabsichtigten Wörter unmittelbar auslesen und sogar in Echtzeit als Text ausgeben lassen. Besonders Patienten, die aufgrund einer schweren Beeinträchtigung ihres Zentralnervensystems beispielsweise nach einem Hirnschlag nicht mehr selbst sprechen können, eröffnet sich damit eine Chance, sich zu verständigen. Das gilt vor allem dann, wenn ihnen aus medizinischen Gründen ohnehin Mikroelektroden auf oder in den Kortex implantiert worden sind, mit deren Hilfe eine KI über ein einfaches Interface ihre Hirnströme auslesen kann. Für eine gesunde Testperson oder gar den Massenmarkt ist eine solche Lösung ungeeignet.

Mit dem Blogeintrag im Juli, der auf diesen wissenschaftlichen Erfolg hinwies, zog sich Facebook allerdings aus dem ehrgeizigen Entwicklungsprojekt für eine nichtinvasive optische Hirn-Computer-Schnittstelle zurück. Stattdessen wolle man künftig ein eher kurzfristiges Entwicklungsziel verfolgen: Armbänder, die mittels Elektromyografie (EMG) die elektrische Muskelaktivität messen. Mit Steuerbefehlen, die über die Nervenbahnen an die Muskeln in Fingern und Händen gehen, sollen Anwender zukünftig intuitiv in der Augmented Reality virtuelle Objekte manipulieren können.

c't kompakt

- Facebook startete sein Entwicklungsprojekt mit dem Ziel eines „Brain-Computer Interface“ für die breite Masse bereits 2017.
- Die Hirnströme bei der Vorstellung von Artikulation lassen sich sowohl in Textform („Brain to text“) als auch akustisch als Sprache ausgeben.
- Die Messung elektrischer Muskelreize ermöglicht es ebenfalls, lautlos artikulierte Worte zu erkennen – und kommt ohne ins Gehirn eingepflanzte Elektroden aus.

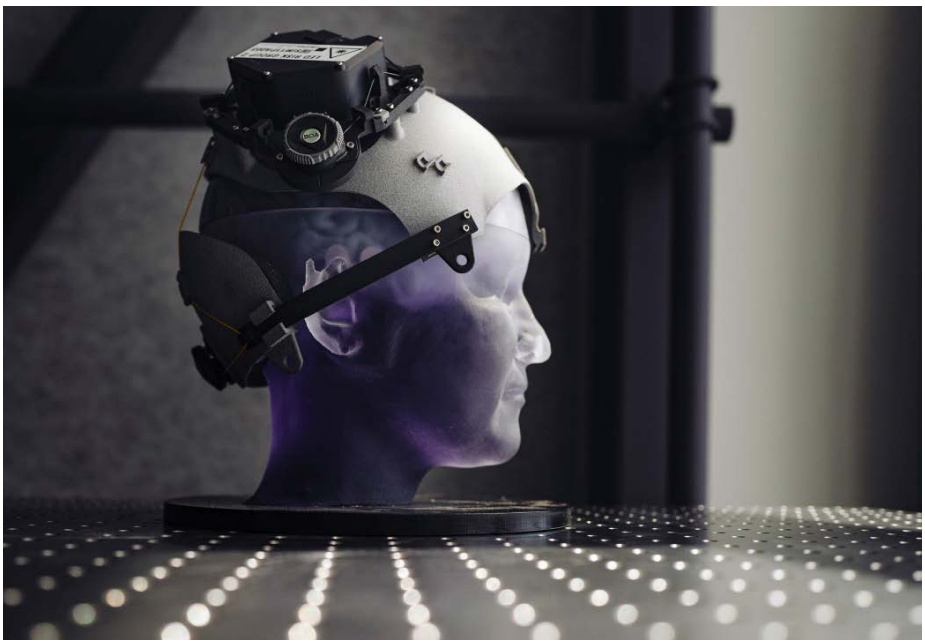

Bild: Facebook Reality Labs

So stellten sich Facebook-Designer eine optische, nichtinvasive Hirn-Computer-Schnittstelle vor. Mit Nahinfrarot beleuchtete Hirnareale sollten ihre Sauerstoffsättigung offenbaren und diese Muster wiederum gedachte Wörter verraten.

Von der einstigen Euphorie gegenüber einem „Brain-Computer Interface“ ist in dem Facebook-Blogeintrag kaum etwas zu spüren. Man hatte sich wohl übernommen. Mark Zuckerberg soll angesichts der Forschungen im Scherz gesagt haben: Das Letzte, was Facebook versuchen solle, sei es, seinen Anwendern buchstäblich den Schädel zu öffnen.

Sprachausgabe in Echtzeit

Genau auf diesem Weg ist die Forschung inzwischen noch einen Schritt weitergekommen. Im September hat ein Team am Bremer CSL die Neurosprachprothese vorgestellt, eine Schnittstelle, die neurale Signale direkt in eine akustische Sprachausgabe übersetzt [2]. Ihr System entwickelten die Bremer in Zusammenarbeit mit dem Department of Neurosurgery an der niederländischen Universität Maastricht und dem Aspen Lab (Advanced Signal Processing in Engineering and Neuroscience) an der Virginia Commonwealth University in Richmond. Die vorgestellte Neurosprachprothese versucht nicht, Wörter zu erkennen, sondern sie setzt allein die Vorstellung der Artikulation direkt akustisch um.

„Unser System funktioniert derzeit mit Tiefenelektroden, die Hirnströme direkt im Inneren des Gehirns aufnehmen und auch mit Elektroden, die unmittelbar auf dem Kortex unterhalb der Hirnschale platziert sind“, berichtet CSL-Direktorin Tanja Schultz. Nur mit einer einfachen EEG-Kappe (Elektroenzephalografie) gelingt das Belauschen der vorgestellten Artikulation von Wörtern eben nicht, weil Schädelknochen, Kopfhaut und Haar sowie Nervenimpulse zu den Kopfmuskeln die genaue Aufnahme der Hirnaktivitäten stören würden.

Der ersten veröffentlichten Studie liegen Versuche mit einer niederländischen Epilepsie-Patientin zugrunde. Ihr waren aus medizinischen Gründen elf Elektrodenbündel mit insgesamt 119 Tiefenelektroden vor allem im linken Frontallappen eingepflanzt worden. In drei Durchgängen baten die Forscher die Patientin, Wörter zu lesen. Im ersten Durchgang sprach sie die Wörter laut aus, im zweiten flüsterte sie sie nahezu lautlos und im dritten sollte sie sich deren Aussprache lediglich vorstellen. In allen Fällen nahmen die Forscher die kurz vorher entstehenden neuralen Signale auf. Tatsächlich zeigte sich, dass die dabei entstehenden Muster vergleichbare Strukturen zeigten, unabhängig davon, ob die Wörter tatsächlich laut gesprochen oder fast stimmlos geflüstert wurden oder ob die Patientin sich deren Aussprache sogar nur vorstellte.

Ein Machine-Learning-Algorithmus lernte, die aufgezeichneten Muster den zugehörigen Lauten zuzuordnen. Konkret ist die niederländische Sprache vergleichbar der deutschen aus etwa 50 bis 60 Lauten aufgebaut. Bei der Zusammenstellung der Übungswörter achteten die Forscher also darauf, dass all diese Laute mehrfach vertreten sind. Heraus kam eine Sammlung aus etwa 75 Wörtern. Dieses Vokabular konnte die trainierte KI schließlich lautweise an eine Sprachausgabe übergeben.

Bild: University of California San Francisco

Die „Brain-to-Text“-Schnittstelle überträgt das EEG eines Multielektroden-Arrays auf dem sensomotorischen Kortex an eine KI. Diese erkennt daraus bestimmte Wörter, die der Patient sich vorstellt.

Erwartungsgemäß hat sich gezeigt, dass der Machine-Learning-Algorithmus für unterschiedliche Probanden individuelle KIs erzeugen muss. Die Lesephase dafür haben die Forscher auf etwa 15 Minuten beschränkt, um die Patienten nicht zu sehr zu belasten. Das automatisierte Training der KI nimmt anschließend noch einmal etwa zehn Minuten in Anspruch.

Kein Tourette-Automat

Wichtig ist Schultz, dass die Neurosprachprothese nicht etwa alles ausplappert, was das Hirn so an sprachlichen Überlegungen entwickelt. Das System vertont nicht irgendwelche Gedanken, sondern es reagiert auf die konkret vorgestellte Artikulation. Es überträgt die dabei entstehenden neuralen Signale in eine lautweise Sprachausgabe und gibt diese nach etwa 400 Millisekunden hörbar aus. Der Mensch hört das Ergebnis seiner Artikulation unmittelbar.

Das System ist nicht nur sehr schnell, es verzichtet auch auf die konkrete Worterkennung, indem es lediglich Laute erkennt und ausgibt. Die Forscher sind nicht den Umweg über eine Worterkennung und eine anschließende künstlich erzeugte Sprachausgabe gegangen. Ihr System protokolliert also auch nicht die gesprochenen oder vorgestellten Sätze des Probanden, es gibt sie lediglich hörbar wieder. Das ist ein qualitativer Unterschied zum seinerzeit von Facebook geplanten „Brain-Computer Interface“. Dafür aber funktioniert das System ungeheuer schnell, tatsächlich in Echtzeit, sodass die künstliche Sprachausgabe erklingt, während der Patient eigentlich erwartet, die eigene Stimme zu hören.

Abgesehen von der geöffneten Schädelplatte hat diese Neurosprachprothese heute allerdings noch einen gravierenden Nachteil: Die Sprachausgabe ist bisher nahezu unverständlich. Das Problem ist, dass die Forscher noch kein System menschlicher Lautbildung implementiert haben. Mit der Umstellung auf komplexere, tiefere neuronale Netze und einen anspruchsvolleren Trainingsalgorithmus hoffen sie, aus den individuell gemessenen neuralen Signalen klarere Wörter erzeugen zu können.

Die eigene Sprache hören und verbessern

Ein zweiter Effekt ergibt sich durch den Patienten im Audiofeedback. Denn der hört die Stimme zu seinen Vorstellungen

wie in einem Regelkreis. „Der Mensch hört sich auch im normalen Leben selbst reden und korrigiert seine Aussprache dabei permanent“, erläutert Schultz. Das merke man besonders, wenn ein Sprecher sich nicht selbst hören kann, bei einem lauten Konzert zum Beispiel. In einem solchen Fall klingt die Sprache schnell ungenau. Sobald sich der Sprecher in ruhigerer Umgebung selbst hören kann, gleicht er Fehler wieder aus. In Versuchen habe sich bereits gezeigt, dass die Patienten mit der Neurosprachprothese sich anpassen und mit der Zeit verständlichere Sprachausgaben erzeugen können.

Tanja Schultz geht nicht davon aus, dass sich gesunde Testpersonen freiwillig Tiefenelektroden ins Gehirn pflanzen lassen, um eine künstliche Sprachausgabe ausprobieren zu können. Das hielte sie auch für unverantwortlich. Für die Studien sind die Forscher daher immer auf Patienten angewiesen, die aus medizinischen Gründen auf neurochirurgische Eingriffe und sehr genaue EEG-Aufnahmen unter der Hirnschale, sogenanntes intrakranielles EEG, angewiesen sind.

In künftigen Studien wollen die Forscher versuchen, Menschen, die am Locked-in-Syndrom leiden, wieder eine Stimme zu geben. Naturgemäß ist die Zusammenarbeit mit diesen Patienten schwierig, da nur einige von ihnen überhaupt eine Rückmeldung etwa über ein Zwinkern oder über Pupillenbewegungen geben können. Bei degenerativen Erkrankungen des motorischen Nervensystems wäre es auch denkbar, dass den Patienten zukünftig eine Neurosprachprothese angepasst wird, solange sie noch sprechen können.

Bild: CSL

Sensoren an der Hautoberfläche messen die elektrischen Muskelaktivitäten der Sprechwerkzeuge. Auch daraus lassen sich mittels einer KI die beabsichtigten Wörter ermitteln.

Lippenleser für alle

Eine Alternative für gesunde Anwender, für die der tiefe Einblick in ihre Hirnströme unangemessen wäre, ist die Messung der elektrischen Muskelaktivität (Elektromyografie, EMG) beim lautlosen Sprechen. Bereits 2010 haben Forscher am CSL diese Technik erstmals auf der CeBIT demonstriert und seitdem verfeinert. Der Gedanke hinter diesem technischen Ansatz ist, dass Lippen, Zunge und Stimmbänder die Sprache artikulieren. Die elektrischen Impulse der dazu erforderlichen Muskelaktivität lassen sich durch Elektroden auf der Haut messen.

Auch bei dieser Technik lernt eine KI, aus den abgeleiteten Signalen die angestrebten Wörter zu erkennen. Diese kann das System der Bremer anschließend textlich weiterverarbeiten oder auch akustisch ausgeben. Das ist nicht nur eine Technik für Menschen, die durch Krankheit oder Unfall ihre Stimme verloren haben.

Das technische Lippenlesen könnte auch gesunde Menschen unterstützen, die in der Umgebung anderer lieber lautlos telefonieren wollen; sei es, um die Mitmenschen nicht zu stören oder auch um vertrauliche Informationen nicht laut vernehmlich auszusprechen. Dadurch, dass der Anwender bei dieser Technik keine Mikrofone einsetzt, kann er auch in lauter Umgebung sprechen, ohne dass der Lärm beim Empfänger ankommt und die Verständigung stört. Letztlich lässt sich die unmittelbare Umsetzung in eine Textversion auch für eine weitergehende Textverarbeitung nutzen, beispielsweise zur Protokollierung oder für eine simultane Übersetzung.

Vokabular über 2000 Wörter

Bereits 2016 konnten die Forscher belegen, dass sie mit dieser Technik selbst große Vokabulare von mehr als 2000 Wörtern mit einer Wortfehlerrate von unter 20 Prozent erkennen können [3]. Durch die Reduzierung des Wortschatzes auf etwas mehr als 100 Wörter konnten sie die Wortfehlerrate auf 3,5 Prozent senken.

In einer jüngsten Arbeit konnten sie die „EMG-to-Speech“-Ausgabe nahezu in Echtzeit verwirklichen [4]. Damit wird es auch möglich, ähnlich wie bei der Neurosprachprothese, die Studienteilnehmer in einen unmittelbaren Regelkreis einzubinden. Sie sprechen lautlos, ihre elektrischen Muskelaktivitäten werden registriert, die unausgesprochenen Wörter erkannt und eine künstliche Sprachausgabe auf den Kopfhörer des Sprechers gespielt. Die Hoffnung ist, dass sich die Erkennungsrate der lautlos gesprochenen Wörter durch das Training im geschlossenen Regelkreis noch erheblich steigern lässt.

Die Ergebnisse dieser Versuche sind jedoch nicht eindeutig. Es scheint, dass der Closed Loop bei manchen Versuchspersonen zu einer besseren Spracherkennung führt und bei anderen wiederum gar nicht. Derzeit grübeln die Forscher noch, wie sich dieser Unterschied erklären lässt.

In jedem Fall könnte diese Technik zur Erkennung lautloser Worte für den alltäglichen Einsatz interessant sein. Allerdings ist eine breite Akzeptanz bei den Anwendern wohl erst zu erwarten, wenn weder operativ eingesetzte Elektroden noch klobige, das Gesicht entstellende Aufkleber für die Spracherkennung erforderlich sind. Nach den Informatikern könnten sich nun auch einmal Designer mit der Gestaltung der Sensoren auf der Gesichtshaut beschäftigen. *(agr@ct.de)* c't

Literatur

[1] David A. Moses et al.; Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria; The New England Journal of Medicine; 15. Juli 2021

[2] Miguel Angrick et al.; Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity; Nature Communications Biology; 23 September 2021

[3] Dissertation von Matthias Janke; EMG-to-Speech: Direct Generation of Speech from Facial Electromyographic Signals; Karlsruher Institut für Technologie; 2016

[4] Dissertation von Lorenz Diener; The Impact of Audible Feedback on EMG-to-Speech Conversion; Universität Bremen; 2021

Weitere Infos: ***ct.de/y779***