Agentur / Ressort:	Wissen		
Veröffentlichungsdatum:	23.06.2019	**Agentur / Herkunft**	kein Eintrag
Text ID:	315592520		
Sperrstatus:	frei	**Priorität:**	kein Eintrag
Notiz:	kein Eintrag		

NZZ am Sonntag Wissen 23.06.2019 / Wissen / Seite 46

Kuck mal, wer da spricht

Computeringenieure haben eine intelligente Software entwickelt, die einzig aus der Stimme eines Menschen ein Phantombild seines Gesichts erstellt. Von Patrick Imhasly?

Das haben Sie bestimmt auch schon erlebt: Sie vernehmen die Stimme eines Ihnen unbekannten Menschen und stellen sich dann vor, wie dieser aussieht. Wenn wir jemanden sprechen hören, durchforstet das Gehirn seine Erinnerungen an bestimmte Gesichter und holt daraus jenes Bild ins Bewusstsein, das am ehesten zu der wahrgenommenen Stimme passt.

Was das Gehirn tut, könnte eine Software vielleicht auch, dachten sich wohl ein paar Tüftler vom Computer Science & Artificial Intelligence Lab am Massachusetts Institute of Technology (MIT) in Boston. Sie stellten sich die Frage: «Was kann man von der Art und Weise, wie eine Person spricht, über ihr Aussehen ableiten?» Erstaunlich viel, wie jetzt eine Studie zeigt, die das Team um den Computeringenieur Tae-Hyun Oh noch vor der offiziellen Publikation auf der Forscherplattform arxiv.org zugänglich gemacht hat. Mit der Methode des sogenannten Deep Learning ist es ihnen gelungen, aus lediglich sechs Sekunden langen Aufnahmen der Stimme von Menschen Phantombilder zu erstellen. Diese geben recht präzise Alter, Geschlecht und ethnischen Hintergrund, aber selbst konkrete Merkmale einer Person wie die Form des Gesichts oder die Struktur der Nase wieder.

Was das Verfahren nicht kann: Es liefert keine Bilder von spezifischen Individuen und erlaubt deshalb nicht, die wahre Identität eines Menschen allein anhand seiner Stimme zu enthüllen. Das hat damit zu tun, dass die Software sprachliche Merkmale von sehr vielen Menschen mit ihren Gesichtszügen in Verbindung bringt und auf dieser Basis im Prinzip durchschnittlich aussehende Gesichter produziert. «Der Ansatz ist eine Spielerei, aber sehr plausibel und höchst interessant», erklärt Volker Dellwo vom Institut für Computerlinguistik der Universität Zürich, der an der Studie nicht beteiligt war.

«Man wird nie so weit kommen, mit einer solchen Methode Menschen individuell erkennen zu können», erklärt der Experte für Phonetik. Aus Sprache erstellte Phantombilder seien deshalb vor Gericht nicht einsetzbar, etwa wenn es darum geht, einen Erpresser zu identifizieren, der seine Forderungen auf ein Aufnahmegerät spricht. «Den Fahndern der Polizei könnten solche Bilder aber durchaus wertvolle Hinweise liefern», sagt Dellwo. «Man

muss einfach wissen, wo die Tücken und die Fehlerquellen liegen.» Was Tae-Hyun Oh und seine Kollegen mit ihrem Verfahren namens «Speech2Face» wirklich vorhaben und ob bereits Sicherheitsbehörden daran interessiert sind – dazu wollten sie sich auf Anfrage nicht äussern.

Stimme und Sprechweise verraten bereits in der klassischen phonetischen Analyse mehr über einen Menschen, als manch einem lieb ist. Ob Mann oder Frau, Kind oder Greis: Die Stimme enthüllt das Geschlecht und bis zu einem gewissen Grad das Alter eines Sprechers. Der Dialekt und der Akzent liefern Hinweise auf die geografische Herkunft und das soziale Umfeld, in dem jemand aufgewachsen ist. Auch der Gemütszustand eines Menschen spiegelt sich in seiner Stimme: Ist jemand nervös und angespannt, tönt diese eher gepresst. «Über den Charakter eines Menschen kann die Stimme aber nichts aussagen», sagt Volker Dellwo.

Niemand versteht, wie es geht

Damit die Software der MIT-Forscher aus unbekannten Stimmproben Phantombilder erstellen konnte, musste sie zunächst einmal entsprechend trainiert werden. Dazu fütterten Tae-Hyun Oh und sein Team das System mit Millionen von Youtube-Videosegmenten von mehr als 100000 Menschen. Das Verfahren habe sich selbst kontrolliert, schreiben sie in ihrer Studie: «Es nutzt einfach das natürliche gemeinsame Vorkommen von Sprache und Gesichtern in Videos und braucht keine weiteren Informationen.» Die Software bildet also Korrelationen zwischen Merkmalen der Sprache und anatomischen Eigenschaften des Kopfes wie der Länge des Kiefers, der Form des Mundes oder der Beschaffenheit der Lippe und verfeinert diese Zusammenhänge durch selbstlernende Prozesse immer weiter.

Dieser Ansatz führt zum Beispiel zu einer verblüffend ähnlichen Rekonstruktion des Gesichts des Bond-Darstellers Daniel Craig (siehe Bilder Seite 45) – stellt aber auch eine Art Blackbox dar. «Man kennt den Algorithmus, versteht aber nicht, aufgrund welcher Kriterien er funktioniert», sagt Volker Dellwo. Ein solches System erfasse Zusammenhänge zwischen Sprache und Gesicht in einem multidimensionalen Raum, «letztlich weiss man indessen nicht, wie diese im Einzelnen rational zu interpretieren sind».

Und doch kommt das Verfahren dem, was das menschliche Gehirn leistet, nahe. «Auch unser Gehirn kennt solche Verknüpfungen», erklärt Katharina von Kriegstein von der Technischen Universität Dresden. «Es weiss, dass sich die Stimme von jemandem mit einer breiten Nase anders anhört, wobei das Gehirn mit diesen Informationen wahrscheinlich viel selektiver umgeht.» Die Neurobiologin beschäftigt sich seit langem mit der Frage, wie das Gehirn Gesicht und Stimme in Beziehung setzt. Dabei hat sie festgestellt, dass die Areale für die Gesichts- und die Stimmerkennung im Gehirn direkt miteinander verbunden sind und sich extrem schnell untereinander austauschen.

Das Ergebnis dieser Kooperation: Wenn wir eine Person nur sprechen hören, aktiviert unser Gehirn die gelernten Assoziationen an das Gesicht, was in einer Art positiven Rückkopplung dazu führt, dass die Stimmerkennung geschärft wird. Im Alltag kann das helfen, vertraute Menschen schnell zu erkennen – selbst

unter vielen Leuten im Supermarkt. «Weniger klar ist, wie das Gehirn reagiert, wenn es eine vollkommen unbekannte Stimme wahrnimmt», sagt Katharina von Kriegstein. «Vermutlich simuliert das Gehirn zu einer bestimmten Stimme aus seinen vielen Erinnerungen den Prototyp eines Menschen.» Es scheint, als leiste «Speech2Face» etwas Ähnliches.