



2

AI: Science
over Fiction

healthAI – Wie Audio Intelligence die Diagnostik und Therapie im Gesundheitswesen revolutionieren kann

Faktenpapier

Aus der Serie: AI: Science over Fiction

www.bitkom.org

bitkom

Herausgeber

Bitkom
Bundesverband Informationswirtschaft,
Telekommunikation und neue Medien e. V.
Albrechtstraße 10 | 10117 Berlin
T 030 27576-0
bitkom@bitkom.org
www.bitkom.org

Verantwortliches Bitkom-Gremium

AK Artificial Intelligence

Projektleitung

Dr. Nabil Alsabah | Bitkom e. V.

Autorin

Dagmar Schuller | audEERING GmbH

Lektorat

Linda van Rennings | Bitkom e. V.

Satz & Layout

Katrin Krause | Bitkom e. V.

Titelbild

© Jeremy Bishop | pexels.com

Copyright

Bitkom 2019

Diese Publikation stellt eine allgemeine unverbindliche Information dar. Die Inhalte spiegeln die Auffassung im Bitkom zum Zeitpunkt der Veröffentlichung wider. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen. Eine Verwendung liegt daher in der eigenen Verantwortung des Lesers. Jegliche Haftung wird ausgeschlossen. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen beim Bitkom.

1 Die KI ist über die automatische Erkennung gesprochener Inhalte weit hinaus

Mithilfe von Methoden der intelligenten Sprachanalyse kann Künstliche Intelligenz verstehen, wie und unter welchen akustischen Bedingungen der Mensch etwas sagt. Denn die Sprache eines Menschen verrät weitaus mehr als nur den Inhalt des Gesagten. Dank Verfahren des maschinellen Lernens in Kombination mit entsprechenden Daten und Modellen können aus dem Audiosignal neben klassischen demographischen Merkmalen wie Geschlecht oder Alter auch menschliche Emotionen, Persönlichkeitsmerkmale und Gesundheitszustände erkannt werden. Über die akustische Szenenerkennung mittels KI können auch soziale Aspekte und Verhaltensweisen in die Bewertung und Kommunikation mit einfließen. Bereiche, die in der zwischenmenschlichen Kommunikation eine Schlüsselrolle spielen, sind in der alltäglichen Mensch-Maschine-Interaktion bisher jedoch nur rudimentär integriert.

Das so genannte Affective Computing ist ein Spezialgebiet der KI. Diese Wissenschaft beschäftigt sich mit der Analyse und Entwicklung von Verfahren, Systemen und Geräten, die menschliche Affekte erkennen und interpretieren sowie diese verarbeiten und stimulieren können. Es handelt sich dabei um einen interdisziplinären Bereich, der neben Informatik auch die Psychologie sowie Kognitionswissenschaften und Teile der Medizin umfasst. Für Audio Intelligence wird dies auch noch weiter um beispielsweise Sprachwissenschaften und Signalverarbeitung (Elektrotechnik) ergänzt.

Mit entsprechenden Verfahren und Modellen, die sich auf Grundlage dieser Wissenschaft entwickelt haben, können umfassende Informationen aus dem Audiosignal abgeleitet werden. Dazu gehören unter anderem Gesundheitszustände, Emotionen und akustische Umgebungseinflüsse. Für einzelne Bereiche liegen die Erkennungsraten der KI-Algorithmen bereits auf menschlichem Niveau, teilweise sogar bereits darüber. Die Verfahren befähigen die Maschinen dazu, den individuellen Zustand eines Menschen zu erkennen, zu bewerten und darauf angepasst und »intelligent« zu reagieren. Damit schafft die Technologie die Grundlage für eine neue Generation von angewandter KI: der Emotionalen Künstlichen Intelligenz. Im Gesundheitssektor übertrifft diese in einigen Anwendungsbereichen aktuell sogar schon die menschliche Leistung, was künftig noch weiter zunehmen wird. In keinem anderen Bereich wird die disruptive Kraft der Technologie so deutlich wie in der Gesundheitsbranche.

»Affective Computing ist ein interdisziplinäres Forschungsgebiet. Es verfolgt das Ziel, KI-Systeme zu entwickeln, die Emotionen verstehen können.«



2 Dank KI wird Sprechen zum wertvollen Biomarker in der Medizin

Zu sprechen und sich zu artikulieren ist eine der komplexesten Funktionen des menschlichen Körpers. Die Artikulationsmuskulatur ist vielfältig aufgebaut. Neben Zunge und Zwerchfell sind viele Muskeln und Körperteile am Sprechen beteiligt. Jede Stimme besitzt kaum veränderbare persönliche und objektive Sprachmerkmale wie Tonlage, Stimmklang, Rhythmus, Betonung oder Melodie. Das gilt vollkommen unabhängig von der Landessprache und dem Inhalt.

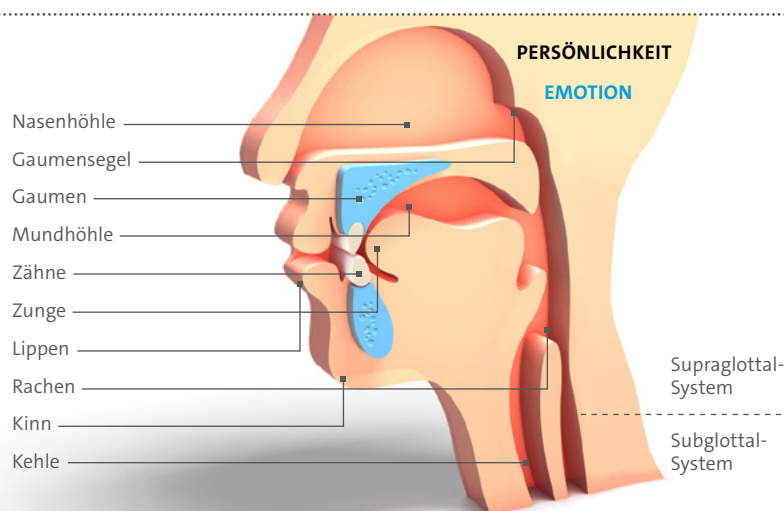


Abbildung 1: Es gibt nur ein Stimmproduktionssystem, aber viele Zustände und Eigenschaften, die Auskunft über unseren gesundheitlichen Zustand geben

Die Sprachanalyse mit KI-Methoden macht Sprachausfälle, Stimmungsschwankungen und Stresszustände in der Stimme sichtbar. Besonders im Bereich der neurokognitiven und neurodegenerativen Krankheiten wie Parkinson, Alzheimer, Autismus oder psychologischen Erkrankungen wie Burnout oder Depression können hier bedeutende Erkenntnisse schnell, effizient und ohne invasive Eingriffe mit mittlerweile robusten Erkennungsraten gewonnen werden. So wird Sprache zum bedeutenden Biomarker in der Medizin und kann in Kombination mit anderen Merkmalen oder Sensordaten zu neuen Erkenntnissen und verbesserten Diagnose- und Therapiemöglichkeiten führen.

3 Intelligente Sprachanalyse erkennt kleinste Veränderungen im Stimmbild von Parkinson-Patienten

Parkinson ist die zweithäufigste Nervenkrankheit der Welt. Allein in Deutschland leiden 400.000 Menschen daran. Die Tendenz ist aufgrund der demografischen Entwicklung weiter steigend. Zittern und verlangsamte Bewegungen galten bisher als alleinige Symptome, die Parkinson erkennen ließen. Die Folge: Eine späte Diagnose der unheilbaren Krankheit und nur geringe Möglichkeiten zur Therapie.

Doch bevor das bekannte Zittern ausbricht, lassen sich bereits Veränderungen im Stimmbild feststellen. Die intelligente Sprachanalyse, entwickelt von audEERING, kann automatisch und mit hoher Genauigkeit motorische Defizite in der Artikulationsmuskulatur erkennen. Die dabei festgestellte Veränderung des Sprachsignals ist für den Menschen nicht wahrnehmbar. Die Analyse kürzester Laute wie »Pah, Tah, Kah« reicht hier aus. Denn mit der Mischung aus Vokalen und Konsonanten kann der Algorithmus aufgrund vorheriger erlernter Erfahrungen die Wahrscheinlichkeit bestimmen, mit der der Patient von Parkinson betroffen ist.

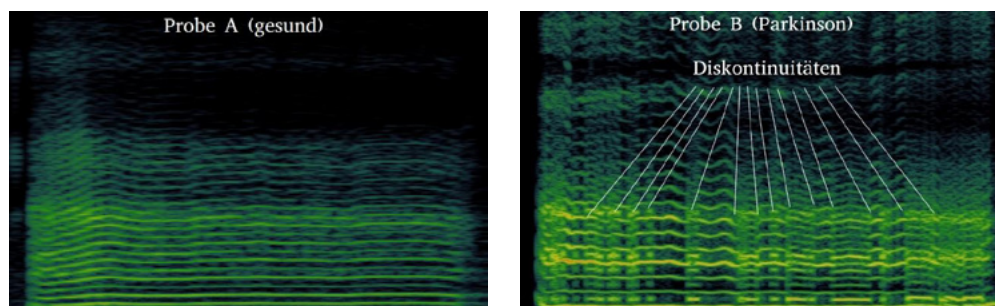


Abbildung 2: Mithilfe von Spektrogrammen lassen sich die Anzeichen einer beginnenden Parkinson-Erkrankung deutlich erkennen: Die Übergänge zwischen den Vokalen und Konsonanten verschwimmen bei dem kranken Patienten (Probe B) deutlich, da er weniger Kontrolle über seine Artikulationsmuskulatur hat

Gemeinsam mit dem britischen Unternehmen emteq hat audEERING diese Technologie zur automatischen Früherkennung von Parkinson-Symptomen bereits in die Anwendung gebracht und durchläuft gerade eine klinische Studie. Die Sprachanalyse-Technologie wurde dabei in emteqs smarte Wearables integriert. Sie ermöglicht durch die kontinuierliche Beobachtung der Symptome eine individuell auf den Patienten zugeschnittene Behandlung. Für die Beobachtung von Parkinson-Symptomen hat emteq eine smarte Brille entwickelt. Sie zeichnet die wichtigsten physischen Indikatoren der Krankheit – wie Veränderungen von Gesichtsausdruck und Bewegungen – auf. Die installierte Software interpretiert diese Aufzeichnungen. Durch die Integration beider Technologien können Mediziner wertvolle nachvollziehbare Informationen für die zuverlässige Früherkennung von Parkinson-Symptomen gewinnen.

4 Wenige Sekunden gesprochener Sprache verraten mehr als 50 Sprecherzustände und Emotionen

Für die Erkennung von Parkinson reicht es, Veränderungen im Stimmbild zu analysieren. Doch für die verbesserte Therapie oder auch Diagnose anderer Krankheiten, wie Depression oder Autismus, sind Informationen über die emotionalen Zustände des Sprechers ausschlaggebend. Auch hier können Verfahren des Affective Computing im Audibereich anhand weniger Sekunden aus dem Sprach-/Audiosignal die relevanten Informationen liefern. Dazu gehören nicht nur das Alter und Geschlecht des Sprechers, sondern auch Emotionen wie Freude, Angst, Trauer oder Wut oder Gesundheitszustände wie Schnupfen oder Husten. Darüber hinaus können diese Verfahren auch die Persönlichkeitsmerkmale sowie den akustischen und sozialen Kontext ermitteln.

Die Auseinandersetzung der KI mit menschlichen Emotionen ist unerlässlich. Denn Emotionen beeinflussen unsere Assoziationen, unser Abstraktionsvermögen und unsere Intuition. Emotionen wirken sich auf unser Wohlergehen aus. Sie lenken unsere Aufmerksamkeit. Und sie beeinflussen unsere Entscheidungsfindung.

Der amerikanische Psychologe Paul Ekman hat sieben Grundemotionen definiert: Freude, Wut, Ekel, Angst, Verachtung, Traurigkeit und Überraschung. Mithilfe einer mehrdimensionalen Matrix kann die KI-basierte Audioanalyse den emotionalen Zustand des Sprechers einer Grundemotion zuordnen und gleichzeitig auch andere Emotionseinflüsse darstellen. Dabei kommen unter anderem die drei Dimensionen Arousal, Valenz und Dominanz zur Anwendung. Arousal schätzt ab, wie erregt der Sprecher ist. Valenz leitet die Attraktivität der vorliegenden Situation für den Sprecher aus der Art und wie Weise ab, wie er gerade spricht. Sie kann entweder positiv oder negativ sein. Eine positive Valenz deutet auf eine wünschenswerte Situation hin. Eine negative Valenz hingegen lässt auf eine unerwünschte Situation rückschließen. Dominanz misst den Nachdruck ab, mit dem der Sprecher redet. Die Lautstärke ist unter anderem ein guter Indikator für die Dominanz des Sprechers.

»Der Algorithmus verortet den emotionalen Zustand des Sprechers auf einer dreidimensionalen Matrix.«

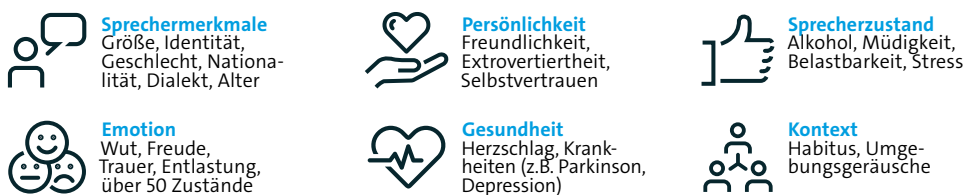


Abbildung 3: Das große Potenzial der intelligenten Sprachanalyse liegt in dem Erfassen und Bewerten von menschlichen Emotionen und (Sprecher-)Zuständen sowie der Prädiktion

Mithilfe dieses mehrdimensionalen Modells kann die Technologie in Echtzeit die Emotionen in der Stimme erkennen und dazu die entsprechenden Wahrscheinlichkeiten und Konfidenzen liefern, um die bestmögliche Bewertung und Entscheidungsgrundlage zu geben. Angst ist z. B. durch hohes Arousal, negative Valenz und niedrige Dominanz gekennzeichnet. Grundemotionen können nur im Zusammenspiel aller Dimensionen richtig klassifiziert werden. Ein hohes Arousal kann beispielsweise sowohl auf Freude als auch Ärger hindeuten. Wird dieser Faktor aber mit der Valenz kombiniert, lassen sich beide Emotionen klar unterscheiden.

5 Tiefe neuronale Netze lernen anhand annotierter Sprachdaten, menschliche Emotionen und Zustände zu erkennen

Bevor die KI Emotionen in der Stimme des Sprechers erkennen, bewerten und die richtige Entscheidung treffen kann, muss sie zuerst mit sogenannten annotierten Sprachdaten trainiert werden. Unter Annotation versteht man die korrekte Zuordnung und Bewertung, das »Labelling« der Daten, was in Folge ausschlaggebend für die Erkennungsleistung des Modells ist. Die Annotation kann durch menschliche Annotatoren erfolgen. Mittlerweile können aber auch für diesen Bereich bereits in Teilen maschinelle Verfahren angewendet werden. audEERING trainiert das Modell mit den aufbereiteten und klassifizierten (menschlichen) Audiodaten. So lernt der Algorithmus, wie sich jemand anhört, der traurig, glücklich oder wütend ist. Für das Trainieren der Modelle werden überwiegend authentische Sprachdaten, die in echten Lebenssituationen entstanden sind, verwendet, um möglichst lebensnahe Situationen und Zustände erfassen zu können. Diese Sprachdaten können z. B. im Rahmen von Forschungsprojekten oder initiierten Aufnahmen entstehen. Ebenso können Audiodaten aus öffentlichen Quellen (beispielsweise Social Media), die mit bestimmten Tags versehen sind, nützlich für die Erstellung von Grundmodellen sein.

6 Das Potenzial der intelligenten Sprachanalyse schöpft man durch hochwertige Daten aus

Affective Computing im Audibereich hat ein enormes Potenzial in vielen Bereichen. Es kann die Kundenbetreuung in Unternehmen unterstützen und ergänzen. Der Mensch-Maschine-Interaktion eröffnet es durch maschinelle Empathie neue Möglichkeiten. So ist es in der Lage, die frühe Diagnose von Krankheiten zu beschleunigen und Therapien zu ergänzen und zu verbessern. Vor allem im medizinischen Bereich erwarten wir große Sprünge. Intelligente Audioanalyse wird in nicht allzu ferner Zukunft zum Standardrepertoire der medizinischen Diagnostik und Telemedizin gehören. Um dies bestmöglich umzusetzen und das Potenzial von KI zu nutzen, müssen die Algorithmen mit ausreichend Daten versorgt werden. Nur so kann man die Erkennungsraten und die Robustheit verbessern.

Wir müssen in Europa Datenschutzhürden überwinden, ohne hierbei den Schutz der Individuen zu beeinträchtigen. Gleichzeitig müssen wir aber die Möglichkeiten im Sinne des Menschen optimal nutzen und anwenden. Wir brauchen einen umfassenden gesellschaftspolitischen Wandel: Weg von der Angst des »gläsernen Menschen« durch KI, hin zu Optimismus und der realistischen Einschätzung der Möglichkeiten, die sich für jeden einzelnen dadurch bieten. Sei es in der täglichen Kommunikation mit Maschinen, sei es durch Anwendungen im Gesundheitsbereich oder in anderen Anwendungsfeldern.

Transparenz und Erklärbarkeit ist hier immanent wichtig, um die Entscheidungen der Systeme nachvollziehen zu können. Die Forschung im Bereich Explainable AI und Transparent AI stimmt einen optimistisch. Sowohl Wirtschaft als auch Politik müssen sich aktiv damit auseinandersetzen, wie sie das Vertrauen der Menschen in die Technologie stärken können. Denn in der öffentlichen Debatte um KI überwiegt die Angst. Diesem Zustand gilt es durch eine sachliche Informationspolitik entgegenzuwirken. Wir brauchen eine innovative und zeitgemäße Bildungspolitik, welche dies ausreichend reflektiert und umsetzt.

»Wir können das Potenzial von Affective Computing nur dann schöpfen, wenn wir genug Trainingsdaten bereitstellen.«

Autorin



Dagmar Schuller

Dagmar Schuller ist CEO und Co-Founder von audEERING.



Bitkom vertritt mehr als 2.700 Unternehmen der digitalen Wirtschaft, davon gut 1.900 Direktmitglieder. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem weltweit führenden Digitalstandort zu machen.

**Bundesverband Informationswirtschaft,
Telekommunikation und neue Medien e.V.**

Albrechtstraße 10
10117 Berlin
T 030 27576-0
F 030 27576-400
bitkom@bitkom.org
www.bitkom.org

bitkom