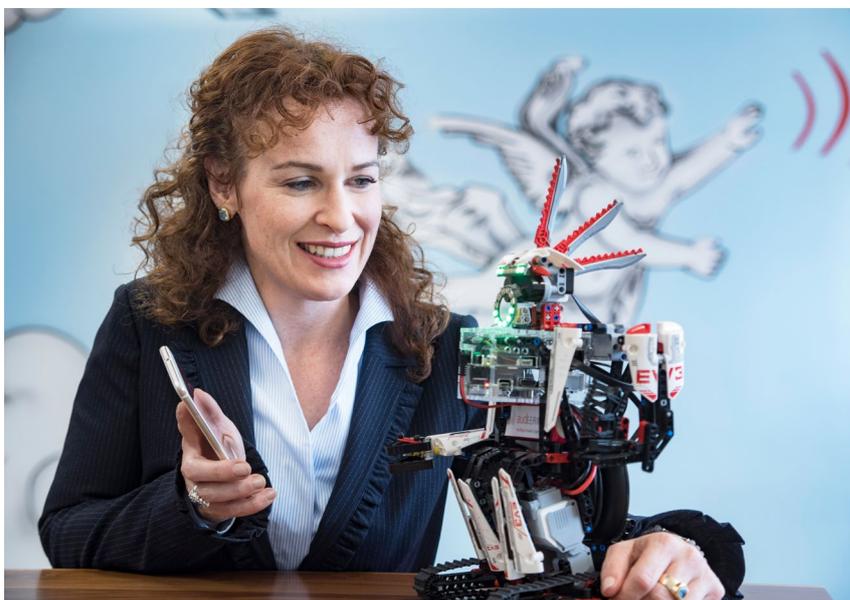


Im Gespräch mit Dagmar Schuller, Geschäftsführerin der audEERING GmbH

Maschinen, die lernen, und Roboter, die denken können... - das Thema Künstliche Intelligenz ist in aller Munde. Und KI wird auch die Möglichkeiten audiologischer Technik erweitern. Doch wohin genau geht die Reise? Was kann Künstliche Intelligenz im Audio-Bereich? Wo sind ihre Grenzen, Chancen, Risiken? – Wir fragten Dagmar Schuller, Geschäftsführerin der audEERING GmbH. Das junge Münchner Unternehmen gilt weltweit als der Innovationstreiber für Audio Intelligenz. Und audEERING ist strategischer Partner des dänischen GN Konzerns.



Martin Schaarschmidt: Frau Schuller, könnten Sie uns audEERING bitte kurz vorstellen. Worum geht es bei Audio Intelligenz?

Dagmar Schuller: Im Kern geht es darum, dass nicht nur zählt, was man sagt, sondern auch, wie man es sagt. Wir lesen mithilfe intelligenter Audio-Analyse Emotionen aus der menschlichen Stimme heraus. Entstanden ist audEERING 2012 als Ausgründung aus der TU München. Unser Ursprung war eine Forschungsgruppe um Professor Björn Schuller. Basis für die Arbeit ist OpenSMILE, eine Software zur automatischen Extraktion von Merkmalen aus Audiosignalen sowie zur Klassifikation von Sprach- und Musiksignalen. Unser System ist heute in der Lage, anhand von wenigen Sekunden Sprachmaterial ungefähr 50 Emotionsklassen zu erkennen. OpenSMILE steht auch als Open Source zur Verfügung. Im Auftrag unserer Kunden – also beispielsweise auch für GN – wird sie dann für konkrete Anwendungen weiterentwickelt.

Martin Schaarschmidt: Das Schlagwort Künstliche Intelligenz ist heute schwer in Mode. Um welche Art Intelligenz geht es da eigentlich?

Dagmar Schuller: Ich würde nicht sagen, dass KI ein Modewort ist. Aktuell erlebt der Begriff eher eine Art Renaissance. KI war schon vor 30 Jahren nichts Ungewöhn-

liches. Nur hatten die neuronalen Netze, mit denen man auch heute beim Machine Learning arbeitet, nur eine Ebene. Bestenfalls gab es zwei Schichten; mehr ging nicht. Heute sind diese neuronalen Netze vielschichtiger.

Unter dem KI-Begriff tummeln sich jetzt eine Reihe neuer Modeworte, beispielsweise Deep Learning. Das heißt eigentlich nur, dass ich ein neuronales Netz mit vielen Schichten habe, sozusagen eine Art Lang- und Kurzzeitgedächtnis abbilden kann. Man muss KI jedoch größer sehen, als es die meisten heute tun. Größer im statistisch-mathematischen und auch im kreativen Sinne. Es geht nicht um einen intelligenten Algorithmus – also nicht um klassisches Machine Learning, das KI überhaupt möglich macht. Es geht vielmehr um ein Zusammenspiel unterschiedlicher Einflussfaktoren. Die ermöglichen ein System, das so ähnlich handelt, wie es ein Mensch tun würde. Aufgrund des von Menschen Gelernten kann dieses System überspitzt ausgedrückt „eigene Gedanken“ entwickeln. Es kann zu Ergebnissen gelangen, zu denen ein Mensch so vielleicht nicht gekommen wäre. Und zwar, weil ich so ein System mit einer Fülle von Daten füttern kann, die ein Mensch gar nicht analysieren könnte. Menschen haben hingegen andere perzeptive Möglichkeiten, die schnellere Schlüsse oder andere Optionen eröffnen. Das ist der große Unterschied zwischen Mensch und KI.

Martin Schaarschmidt: In sehr vielen Branchen hört man derzeit, dass Unternehmen bereits KI einsetzen...

Dagmar Schuller: Meist ist das jedoch gar keine KI, sondern eine statistische Methode. Bestenfalls werden noch zwei Methoden kombiniert. Man muss also genau nachfragen, was dort geschieht. Mit welcher anderen Methode wird die eigene verglichen, um zu sehen, dass man auch das beste Ergebnis hat? Hier trennt sich die Spreu vom Weizen. Alles ist stark Daten-getrieben. Und nur bestimmte Daten mit einer bestimmten Qualität ermöglichen es dem System, in einer bestimmten Qualität zu erkennen.

Das ist nicht anders als bei einem Kind. Sie können ihm Wissen systematisch oder völlig unsystematisch vermitteln. Sie können ihm z. B. systematisch vermitteln: das ist ein Buch, das eine Zeitschrift, das ein Heft. Fehlt diese Systematik, lernt das Kind zwar immer noch was. Aber es weiß dann vielleicht nicht: Buch, Zeitschrift, Heft. Es weiß nur: Das sind alles Sachen, in denen man lesen kann; und sie schauen unterschiedlich aus.

Martin Schaarschmidt: Wie trainieren Sie Ihr System, damit es lernt, Sprecher zu analysieren?

Dagmar Schuller: Geht es z. B. um die Wahrnehmung einer bestimmten Emotion, die in der Stimme mitschwingt, so werden viele Menschen diese Emotion

gleich beurteilen. Sie stellen fest: Das klingt glücklich, verärgert usw. Wir lassen unsere Daten auch von Menschen annotieren bzw. einordnen. Wir haben ein großes Team aus Annotatoren. Aus ihrem Urteil bilden wir ein Mittel. Und anhand dessen wird dann die KI trainiert.

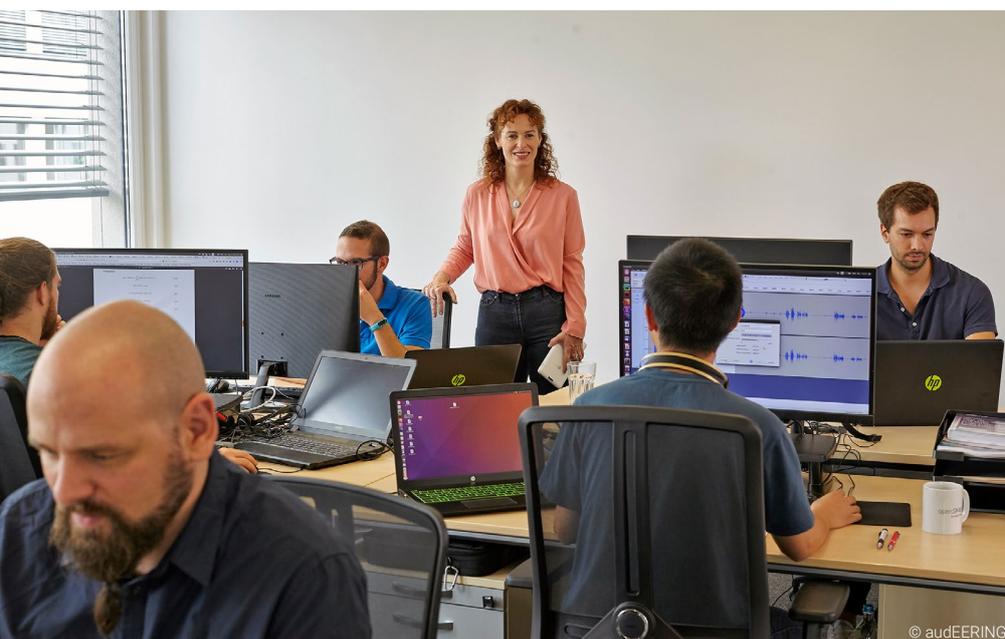
Dieses Systematisieren, das Labeln der Daten, ist extrem wichtig. Nur so kommt man überhaupt zu einer guten Methode. Es wird zwar auch heute schon darüber gesprochen, die Generierung und das Labeln der Daten komplett den Maschinen zu überlassen. Aber dieses sogenannte Automated Machine Learning steckt noch in den Kinderschuhen. Aktuell muss das noch komplett von Menschen übernommen werden, um dann Algorithmen entwickeln zu können, die aus den Daten das Bestmögliche herausholen.

Martin Schaarschmidt: Kann KI tatsächlich besser wahrnehmen als wir? Gleicht sie vielleicht nur aus, dass wir unsere Fähigkeit zur akustischen Wahrnehmung seit Beginn der Industrialisierung sehr vernachlässigt haben?

Dagmar Schuller: Jein. Kommen wir zurück zu den Daten. Auch alle nicht verbalen Äußerungen – ein Zögern oder Lachen, Weinen, Seufzen, Stöhnen – all das muss erst von Menschen annotiert werden, bevor wir es einer Maschine beibringen können. Die lernt dann: Das ist das und das und das. Allerdings hören die Menschen, die für uns annotieren, all diese Details auch genau auf

den Punkt. Sie nehmen alles so genau wahr, wie man es im normalen Leben vielleicht nicht tun würde. Die Maschine hat dann die Möglichkeit, genau dieses Defizit auszugleichen. Dafür muss man ihr aber erst einmal sagen, was das und das ist, damit sie es überhaupt erkennen kann. Hat sie es gelernt, kann sie es besser wahrnehmen als jemand, der nicht genau hinhört. Sie ist dann darauf trainiert z. B. nichtsprachliche Teile der Sprache als gleichwertig wahrzunehmen.

Das Abbilden dieses gesamten Bereichs der Paralinguistik ist ein großer Vorteil unserer Technologie. Wir



sehen die Audiosignale und den gesamten Menschen. Wir sehen nicht nur bestimmte Aspekte wie Tonalität oder Sprechrhythmus. OpenSMILE ist schon heute auf das Erkennen von über 6.000 Charakteristika trainiert. Ein Zögern, eine bestimmte Betonung, Aussprachevarianten, Dialekte... - Es wird nicht nur der Ton gehört, sondern auch die Schwankung, der Verlauf eines Signals oder Atemgeräusche, die zum Beispiel fast komplett nicht tonal sind. Dieses „Ha“ bei einem Erschrecken. Da hört man ja kaum etwas. Aber das System erkennt, wie aufgeregt jemand ist oder wie dominant er spricht. Wir können Persönlichkeitsmerkmale aus der Sprache bzw. aus der Stimme herausfiltern; wie extrovertiert jemand ist, wie empathisch er reagiert... Es sind Dinge, die man sehr vielfältig einsetzen kann.

Redaktion: Etwa im Gesundheitsbereich, in dem Sie stark engagiert sind?

Dagmar Schuller: Das ist ein wichtiges Feld für uns. Hier arbeiten wir an diversen Forschungsprojekten mit. Etwa zur Früherkennung von Depression, Alzheimer, Parkinson, die sich oft schon sehr früh in Sprache bzw. Sprechen niederschlagen – lange bevor sie auf andere Art wahrnehmbar werden. Beim Forschungsprojekt zum Thema Depression etwa geht es um Frühdiagnostik und um Therapie. Ansatz ist hier, dem Patienten ein Audio-Tagebuch an die Hand zu geben. Jeden Tag spricht er in sein Handy, wie der Tag war und wie es ihm ging. Wenn der Patient es wünscht, könnte er das dann an seinen Arzt oder Therapeuten schicken; der erhält dann durch die Analyse ein deutlich komplexeres Bild als das, was er sich einmal in der Woche bei der unmittelbaren Begegnung mit dem Patienten machen kann. Beim Arzt dauern diese Begegnungen ja oft nur fünf bis zehn Minuten.

In einem anderen Projekt geht es um autistische Kinder. Sie haben große Schwierigkeiten, Emotionen anderer richtig zu deuten oder auch eigene Emotionen zu zeigen. Hier haben wir einen Roboter entwickelt, der den Kindern spielerisch hilft, Emotionen zu erlernen – wie eine Art Fremdsprache: Das ist glücklich, traurig, wütend... - Mit einem Roboter fällt den Kindern dieser Lernprozess deutlich leichter als mit einem Menschen. Bei Menschen haben sie oft Barrieren. Mit einem Roboter, der wie ein Spielzeug aussieht, können sie viel leichter interagieren.

Martin Schaarschmidt: Wie ist es bei Alzheimer?

Dagmar Schuller: Hier ist das Projekt, an dem wir mitarbeiten, noch relativ am Anfang. Alzheimer-Diagnostik ist auch für den Neurologen herausfordernd, da es mehrere Varianten gibt. Man diagnostiziert nicht einfach Alzheimer sondern erkennt typische Marker. Betrachtet man Veränderungen in der Stimme, so gibt es bei Alzheimer drei verschiedene Biomarker. Zum einen geht es um klassische Emotionserkennung; Alzheimer-Patienten können von einem Moment auf den anderen emotional ausbrechen. Diese Wutausbrüche sind ganz typisch für Alzheimer. Ein zweiter Marker betrifft eher das Sprechen: Diese Patienten benutzen Worte oft nicht mehr im richtigen Kontext; in einem Satz tauchen plötzlich Worte auf, die da inhaltlich nicht hineinpassen. Dritter Marker ist die Aussprache. Die Färbung von Vokalen und auch von Konsonanten ändert sich. Man spricht langsamer, verzögert häufig. Das ist so eine klassische Anomalie.

Unser System braucht nun für alle drei Marker entsprechende Daten. Idealerweise müssten wir die Daten von Menschen vor der Erkrankung haben und danach. Das ist herausfordernd. Bei Parkinson hingegen haben wir bereits sehr valide Ergebnisse für zwei Klassen. Wir können anhand eines Sprachsamples von unter 15 Sekunden erkennen, ob jemand Parkinson hat oder nicht.

Martin Schaarschmidt: Wie läuft eine solche Diagnose praktisch ab?

Dagmar Schuller: Mit einem klassischen Sprachtest, bei dem Vokale gefiltert werden. Das A hört sich bei diesen Menschen anders an. Bestimmte Testaufgaben können sie nicht lösen – etwa mehrmals kurz hintereinander „pataka pataka pataka“ sagen. Wir sind auch hier noch nicht am Ende. Aber wenn man sich vorstellt, dass das eine Eigendiagnose ermöglicht, für die man nicht erst zu Arzt oder Neurologen muss, ist das schon ein guter Schritt. Bei dem Projekt, das wir zusammen mit einem britischen Unternehmen realisiert haben, liegt die Erkennungsrate heute bei 92 Prozent – bei Verwendung einer Sprachprobe mit 15 Sekunden Länge. Da die sprachlichen Veränderungen bei Parkinson früher auftreten als die motorischen, wird eine frühzeitigere Diagnose möglich. Um noch zuverlässiger zu sein, brauchen wir noch bessere Daten.

Martin Schaarschmidt: Ihre Lösungen sind aber auch in ganz anderen Branchen gefragt?

Dagmar Schuller: Das stimmt, etwa im Call-Center-Bereich: Ein Anrufer meldet sich; denn er hat ein Problem und ist verärgert. Der Agent im Call Center muss ihn an diesem Punkt abholen, und unsere Lösung hilft ihm dabei. Sie misst z. B., wie sich die Emotionalität des Anrufers während des Gesprächs verändert. Das kann im Nachgang zu Trainingszwecken dienen. Die Werte stehen aber auch schon während des Gesprächs in Echtzeit zur Verfügung. Der Call Center Mitarbeiter sieht das als Kurve vor sich auf dem Bildschirm, und er kann dieses unmittelbare Feedback für seine Gesprächsführung nutzen.

Wichtige Rückschlüsse kann Emotionserkennung auch in der Marktforschung liefern. Hier lässt sich z. B. anhand bestimmter Parameter messen, wie gut ein Proband eine Marke oder ein Produkt findet. Traditionell nutzt man hier Befragungen Face-to-Face. Doch es kann vorkommen, dass ein Proband etwas anderes sagt, als

er tatsächlich meint. „Eigentlich ganz gut“ ist vielleicht nur eine höfliche Umschreibung von „nicht gut“

Martin Schaarschmidt: Zu Ihren Kunden gehören aber z. B. auch BMW und Daimler...

Dagmar Schuller: Im Automobilbereich geht es zum einen um die Analyse von Geräuschen im Fahrzeug, etwa von Motorgeräuschen. Gegenstand ist aber auch, was im Auto gesprochen wird. Das ist ein bisschen wie bei K.I.T.T., dem Auto in der TV-Serie Knight Rider. Hier versteht das intelligente Auto seinen Besitzer, den Serien-Helden Michael Knight, perfekt. Es merkt, wenn er müde oder verärgert ist. Es gibt ihm Tipps. Es sagt zum Beispiel „Du bist sehr müde.“ Und bietet an, den Autopiloten zu starten. Heutige Sprachsysteme in Autos können das noch nicht. Doch unser System kann Müdigkeit oder auch Trunkenheit an der Stimme erkennen. Es kann Warnhinweise geben oder auch für Komfort sorgen.

Martin Schaarschmidt: Ihre KI wird für verschiedenste Sprachen genutzt. Wie ist das mit kulturellen Unterschieden? Ich stelle mir z. B. mal ein italienisches Paar vor, das sich streitet. Für jemanden aus Deutschlang klingt das vielleicht dramatischer, als es tatsächlich ist. Berücksichtigt die KI auch das?

Dagmar Schuller: Ja. Wobei der Grundstock der Emotionserkennung unabhängig von einer bestimmten Sprache funktioniert. Die Basis-Emotionen lassen sich ohne eine bestimmte Sprache oder Kultur herleiten. Natürlich ist es immer besser, wenn Sie bestimmte Sprach-Corpora berücksichtigen – etwa für Italienisch, Russisch, Chinesisch. Wir haben so etwas bereits vor Jahren entwickelt und z. B. in der Marktforschung mit der GfK getestet. Man kann die Erkennung verbessern, wenn man den Bezug zu einem bestimmten Sprachkorpus hat. Aber diese Corpora sitzen auf unseren Ur-Emotionen. In diesen Ur-Emotionen sind wir Menschen überall gleich.

In der Steinzeit gab es noch keine Sprache. Das Gegenüber grunzte Sie an. Und in der Art und Weise, wie es grunzte, erkannten Sie, ob Sie laufen müssen oder bleiben können. Das ist wie bei einem knurrenden Hund. Man merkt, ob er spielen will oder gleich zubeißen wird. Der Bezug zum Kulturkreis verbessert das nur noch mal. Chinesische Konnotatoren werden



eher merken, ob ein chinesischer Sprecher mehr oder weniger verärgert ist. Andere erkennen das grundsätzlich auch, können aber nicht so genau abstufen. Bei den streitenden Italienern ebenso. Wir erleben sie als hoch dramatisch. Ein Italiener würde sagen: Alles normal; das ist in fünf Minuten erledigt. Es ist der feine Unterschied – und nichts anderes als eine zusätzliche Modellvariante, die das System trainieren kann.

Martin Schaarschmidt: Haben Sie eigentlich auch mal das Sprechen von Politikern analysiert?

Dagmar Schuller: Tatsächlich haben wir auch daran geforscht – im österreichischen Wahlkampf und auch bei Trump und Clinton. Bei Trump war erstaunlich, dass er in unserer Analyse eine sehr hohe „agreeableness“ erzielte. Die unterbewusste Wirkung, die die Art seines Sprechens auf die Leute hat, war überraschend positiv. Es kann ein Grund sein, warum er die Wahl gewonnen hat

Martin Schaarschmidt: Wo steht Ihr Unternehmen aktuell beim Wettbewerb um KI im Audo-Bereich?

Dagmar Schuller: Wir sind wirklich stolz darauf, der weltweite Innovationsführer auf dem Gebiet der Emotionserkennung aus der Stimme zu sein. Mit unserer Forschung an der TU München haben wir die Basis für die Technologie gelegt, die seither das ein oder andere Mal kopiert wurde. Die Grundlegende Idee der Emotionserkennung aus Sprache stammt von uns, also aus Deutschland und nicht aus den USA, wie viele manchmal denken.

Heute, in der Zusammenarbeit mit Unternehmen, veröffentlichen wir natürlich weniger. Aber früher war das der Fall. Der Begriff Deep Learning etwa ist seit 2012/13 ein viel benutztes Buzzword. Wir waren die allerersten, die 2009 Ansätze von Deep Learning auf Audio-Daten angewandt haben. Wir waren die ersten, die nachgewiesen haben, dass man mit größeren neuronalen Netzen auch bessere Ergebnisse erzielen kann. Überhaupt hat niemand vor uns Emotionserkennung in Sprache gemacht. - Also bahnbrechende Dinge, die wir auch auf Konferenzen und in Journalen vorgestellt haben.

Martin Schaarschmidt: In welchen Punkten heben Sie sich heute von anderen ab?

Dagmar Schuller: Ein entscheidender Unterschied ist, dass unsere Konkurrenz ihre Systeme oft mit Schauspielern trainiert. Wenn die dann vor dem System Ärger spielen, dann ist die Leistung beim Erkennen oft extrem hoch – häufig sogar besser als bei unserem System. Das liegt aber daran, dass diese anderen Systeme nicht auf normale Menschen trainiert sind, die ihre Emotionen ganz normal äußern. Wenn hingegen unserem System eine Emotion vorgespielt wird, heißt das noch lange nicht, dass es diese Emotion auch als solche erkennt. Es ist eben auf Otto Normalverbraucher konditioniert und nicht auf Meryl Streep, die eine Emotion als Schauspielerin ganz anders darstellt. Das sind wesentliche Unterschiede. Wenn jemand verärgert ist, ist er nicht nur verärgert. Vielleicht ist er zugleich frustriert oder traurig. Emotionen haben unterschiedliche Ausprägungen. Sie können miteinander vermischt sein. Und die Maschine muss nun entscheiden, ab welchem Grad sie wie erkennt. Um diese graduellen Applikationen zu erhalten, ist unser System das Beste.

Auch den paralinguistischen Bereich berücksichtigen viele Wettbewerber nicht. Viele Klassifikationssysteme arbeiten nach wie vor textbasiert, während sie nonverbale Äußerungen kaum klassifizieren. Hinzu kommt unsere klassische Erkennung akustischer Szenen, die wir auch bei Gesprächen vorschalten können. Auch das haben die anderen nicht. Laut Gartner Report vom August 2018 zum Thema emotionale AI sind wir in unserem Bereich nicht nur der Anbieter mit dem breitesten Erfahrungsspektrum und dem höchsten Innovationsfaktor, sondern auch der mit den besten Anwendungsmöglichkeiten. Da wurde weltweit analysiert. Wir sind im Gegensatz zu der Konkurrenz auch breiter aufgestellt. Unsere Technologie ist im Gesundheitsbereich, in der Automobilwirtschaft, im Call-Center Bereich sowie in der Marktforschung einsetzbar. Viele konkurrierende Unternehmen sind hingegen auf eine Branche, häufig das Marketing, spezialisiert. Und wir sind ständig auf der Suche nach weiteren Anwendungsbereichen.

Martin Schaarschmidt: Welchen Stellenwert hat für Sie die Zusammenarbeit mit GN?

Dagmar Schuller: Diese Partnerschaft ist sehr, sehr wichtig für uns. Wir haben uns nicht einfach für Jabra und ReSound entschieden, weil es keine anderen Optio-

nen gegeben hätte. Wir hätten klassisches Venture Kapital oder auch andere Partner wählen können. Für unsere Entscheidung gab es drei Gründe: Zum einen ist GN ein toller Hardware-Hersteller; wir haben großen Respekt vor den fundiert entwickelten und hochwertigen Produkten, vor der Art und Weise, wie sie entstehen. Diese Hardware- und unsere Software-Kompetenzen ergänzen sich perfekt. Das passt sozusagen wie Deckel auf Topf.

Der zweite Grund ist, dass wir durch GN die Möglichkeit haben, unsere Software direkt an den Endkonsumenten zu bringen. Das entspricht unserer Vision. Wir sind kein B-to-C-Anbieter. Wir bieten im B-to-B anderen Unternehmen die Möglichkeit, unsere Technologie in ihre Produkte zu integrieren. Und wir suchen dafür weltweit agierende Partner, die unsere eigenen hohen Qualitätsansprüche teilen. Diese Partner erhalten durch uns einen entscheidenden Unterschied innerhalb ihres Wettbewerbs – so wie das aktuell beim Jabra Elite 85h der Fall ist.

Martin Schaarschmidt: Und Punkt drei?

Dagmar Schuller: Das ist mehr ein emotionaler: Bei GN hatten wir von Beginn an das Gefühl, dass sie mit uns gemeinsam etwas schaffen möchten – nicht nur für den Markt, sondern auch für den Endkonsumenten. Ein optimales Produkt mit wirklichem Mehrwert. Es ist eine gemeinsame Vision, für die wir gemeinsam einen großen Schritt in Richtung Zukunft gehen. Und wir dürfen dennoch unsere Freiheit behalten und können weiter innovativ und kreativ sein. Wir haben in GN einen Partner gefunden, der über Innovationen so ähnlich denkt wie wir, der nicht nur die nächsten zwei Jahre im Blick hat.

Martin Schaarschmidt: Sie erwähnten schon Jabra Elite 85h, das erste gemeinsame marktreife Produkt von GN und AudEERING. Wo liegt die Besonderheit?

Dagmar Schuller: Es ist der weltweit intelligenteste Active Noise Cancellation (ANC)-Kopfhörer. Er wurde Anfang des Jahres vorgestellt und ist mittlerweile im Handel. Er erkennt akustische Szenen und passt sich an diese an. Wechselt man etwa von einem lauten Bahnsteig in ein leises Zugabteil, dann stellt er sich nicht nur automatisch ein, sondern auch individuell, so wie man es gerne hätte.

Bei Jabra Elite 85h geht es um die Analyse akustischer Szenen, die adaptiv über die Smart Sound App erfolgt. Aber das ist nur ein erster Schritt. Es soll weitere Produkte mit unserer Technologie geben. Es geht um optimale Hörerlebnisse für unterschiedlichste Situationen, um das Anpassen an Situationen, um Signale, die herausgefiltert oder durchgelassen und optimal wahrgenommen werden. - Nicht einfach nur Noise Cancelling. Je nach Situation wird nur ein bestimmter Teil herausgefiltert oder verstärkt.

Martin Schaarschmidt: Und bei den Hörgeräten?

Dagmar Schuller: Bei Menschen mit Hördefizit liegt der Focus natürlich etwas anders als bei Kopfhörern. Diesen Unterschied sehe ich gerade mit Blick auf die Wahrnehmung von Emotionen, die man nicht sehen kann – etwa am Telefon. Diese emotionale Komponente einzubringen bzw. zu verstärken, könnte den Hörgeräte-Trägern noch einen deutlichen Vorteil bringen. Und Hören ist ja höchst individuell. In meiner Familie gibt es sehr viele Personen mit einem Hördefizit. Aber jeder von ihnen hört anders. Auch zwischen dem weiblichen und dem männlichen Hören gibt es meiner Meinung nach Unterschiede. Und gerade unsere Technologie ermöglicht ein entsprechend hohes Maß an Individualisierung.

Ich gehe davon aus, dass die Hörgeräte der nächsten Generationen die jeweilige akustische Szenerie noch viel besser analysieren werden, als sie das heute schon tun. Diese Medizinprodukte werden sich noch viel besser automatisch auf die jeweilige Umgebung einstellen. Sie werden erkennen, dass ich gerade im Restaurant sitze, dass ich mich unterhalten will. Sie werden Störgeräusche noch besser filtern und sicherstellen, dass ich den Kellner verstehe... Und es wird um Wahrnehmung gehen.

Martin Schaarschmidt: Inwiefern?

Dagmar Schuller: Wie reagiere ich selbst auf die Szenerie, auf die sozialen Gegebenheiten? Wie ist mein aktueller emotionaler Zustand? Geht es mir gut oder eher nicht? - Die KI wird das wahrnehmen und die Technik dementsprechend nachjustieren. In der Folge wird es darum gehen, nicht nur einzelne Szenarien abzubilden, sondern den gesamten Tagesablauf. Die Technik stellt sich jederzeit so ein, dass es mir gut geht und ein Optimum an Outcome sichergestellt ist.

Voraussetzung ist allerdings, dass ich zwei Dinge akzeptiere: Ich muss wissen, dass ein Teil meiner Daten transparent ist und fortlaufend analysiert wird; damit muss ich konform gehen. Und ich muss zum zweiten die Technik so diszipliniert nutzen, dass sie für meine Wünsche trainiert wird. Ich muss zurückmelden, was ich gut oder nicht gut finde. Sonst kann das System nicht auf mich zugeschnitten werden. Je mehr ich mich mit der Technik austausche, umso besser die Möglichkeiten zur Anpassung.

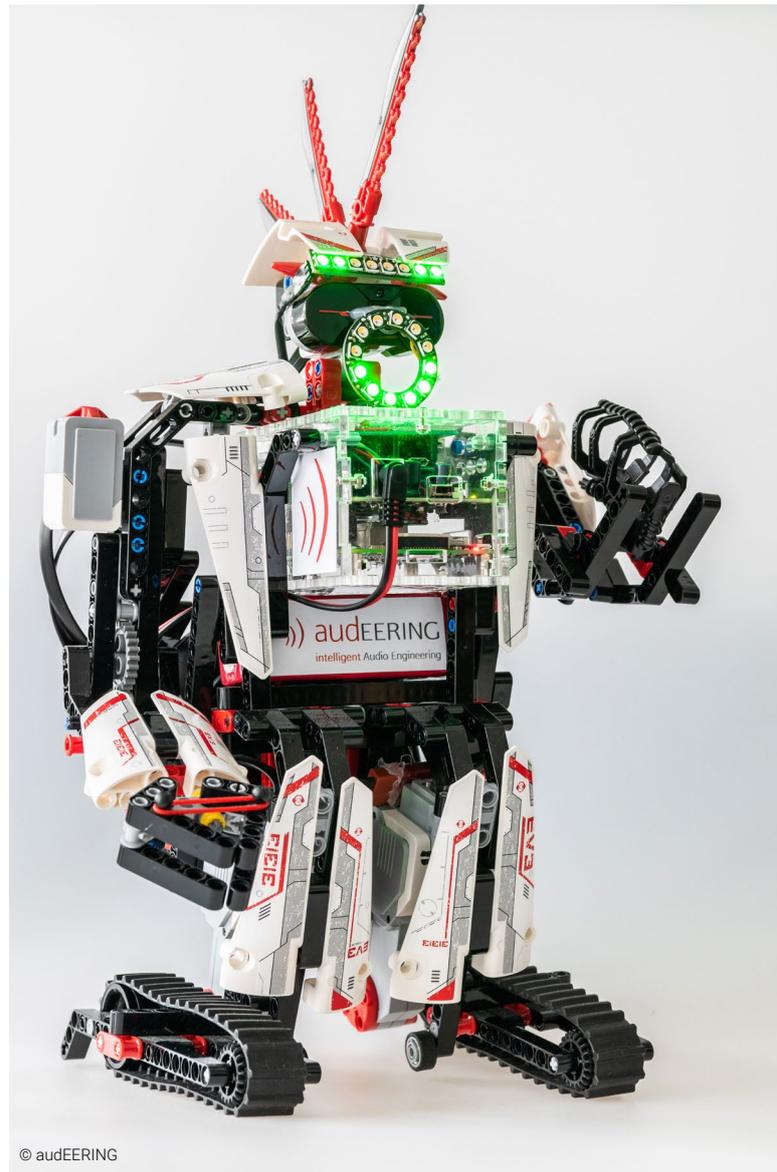
Martin Schaarschmidt: Welchen Stellenwert hat bei Ihrer Arbeit die Sicherheit von Daten?

Dagmar Schuller: Wir sind ein deutsches Unternehmen. Das ist schon anders als bei einem amerikanischen Unternehmen, für das Datensicherheit vielleicht erstmal nicht die Rolle spielt wie hier in Deutschland. Wir ermöglichen z. B. auch, unsere Technik embedded auf einem Gerät zu installieren, ohne dass die Daten in eine Cloud wandern, aus der sie abgegriffen werden könnten. So erhält der Nutzer die Daten bei sich auf seinem Endgerät. Der Nutzer muss auch entscheiden können, ob er seine Daten für seinen Arzt freigeben möchte oder nicht. Er kann die Depressionsdiagnostik auch nur als eine Art Selfmonitoring nutzen. Das ist die Frage nach dem Umgang mit einer solchen Technologie; vor der steht die Gesellschaft als Ganzes.

Martin Schaarschmidt: Mein Eindruck ist, dass Politiker und Politikerinnen oft nicht in der Lage sind, solche neuen technologischen Möglichkeiten adäquat in ihren Entscheidungen abzubilden?

Dagmar Schuller: Sie informieren sich oft auch zu wenig. Ich wurde z. B. von der Leiterin für Digitalisierung im Arbeitsministerium angesprochen. Sie schilderte mir, dass sie in ihrem Ministerium bestimmte Punkte zur Digitalisierung voranbringen müsse. Und sie gab offen zu, sie müsste sich da erst einarbeiten. Wir haben uns mehrfach getroffen, und ich habe das sehr begrüßt. Es kann nicht jeder alles wissen. Vieles ist speziell, und nicht jedem liegt Mathe oder Statistik. Aber es ist wichtig, dass die Politik den Dialog sucht.

Es gibt Leute in der Politik, die sich sehr viel informieren. Natürlich gibt es auch solche, denen PR wichtiger ist als die eigentliche Sache – und entsprechend sind



dann die Ergebnisse. Aber eigentlich sollte sich jeder von uns mit diesen Dingen auseinandersetzen, bevor er sich dafür oder dagegen entscheidet. Das sollten wir schon deshalb tun, weil diese Entwicklungen nicht nur in Deutschland oder in Europa stattfinden. In China und in den USA werden diese Themen auch intensiv bearbeitet. Ich glaube, wir übersehen Großteils, dass die Technologien, die wir nutzen, auch uns nutzen, ohne dass es uns überhaupt bewusst ist. Deswegen ist die offene und transparente Kommunikation so wichtig – ohne Ängste zu schüren. Und die Politik muss den mündigen Bürger sehen – nicht nur die nächste Wahl. Sie muss sich vielfältig beraten lassen, Entwicklungen in der Grundlagenforschung kennen und auch das, was durch die Industrie umgesetzt werden kann. Erst dann kann

sie Dinge angemessen kommunizieren und in Gesetzen regeln – nicht nur, um das Land voranzubringen, sondern auch um bestmögliche Bedingungen für den Bürger zu schaffen.

Martin Schaarschmidt: Mir fallen gleich mehrere Bereiche ein, in denen ich die Nutzung Ihrer Technologie kritisch fände. Was ist z. B. mit Leuten, die Jobs oder Kredite vergeben? Wie ist es mit einer militärischen Nutzung? Wo sind die ethischen Grenzen?

Dagmar Schuller: Da positionieren wir uns ganz klar: Für Anwendungen im militärischen Bereich stehen wir nicht zur Verfügung. Und wir möchten, dass diejenigen Firmen, die unsere Technologie einsetzen, auch transparent gegenüber ihren Kunden kommunizieren, dass sie sie einsetzen.

Andererseits ist mir eine optimistische Sicht auf KI sehr wichtig. Hier in Deutschland tendieren wir immer dazu, solche Entwicklungen sehr negativ zu sehen. In den USA ist es genau umgekehrt. Und in China redet niemand darüber, da wird es einfach gemacht. Es ist eine gesamtgesellschaftliche Problematik, und hier in Deutschland stehen wir uns manchmal ein bisschen selbst im Weg.

Martin Schaarschmidt: Was würden Sie empfehlen?

Dagmar Schuller: Wir brauchen einen transparenten und ausgewogenen Mittelweg. Man kann mit der Technologie unglaublich viel erreichen. Und man muss sich kritisch auseinandersetzen. Man muss abwägen: Was kann ich kriegen und was muss ich dafür geben? Und man muss entscheiden können: Ist es mir das wert oder nicht. Ein Mensch, der chronisch unter Depressionsschüben leidet, wird es sicherlich begrüßen, wenn er eine bessere Therapie mit weniger Medikamenten und weniger Nebenwirkungen bekommen kann. Ob ein gesunder Mensch seine Daten liefert, um am Ende auf zehn Kilometer fünf Minuten schneller zu laufen, ist eine andere Entscheidung. Es hilft nicht, ständig nur zu schimpfen. Und es ist auch nicht gut zu verdrängen, wer ständig Daten von mir einsammelt – wenn ich z. B. auf Facebook like oder Alexa mithört. Es braucht mehr Verständnis für diese Technologien – für ihre negativen und für ihre positiven Seiten. Und die Unternehmen müssen transparent arbeiten. Der Nutzer muss sich entscheiden können. Und es muss ihm klar werden, welche Vorteile

ihm im konkreten Fall versagt bleiben, wenn er sich gegen eine Nutzung entscheidet.

Martin Schaarschmidt: Was, wenn sich die KI irrt?

Dagmar Schuller: Auch wenn ein Mensch einen anderen wahrnimmt – etwa bei einer Polizei-Befragung, dann nimmt er immer gewisse Dinge wahr. Auch das ist ungefähr bzw. gefärbt. Vielleicht hat der Polizist einen schlechten Tag. Der Mensch ist nicht immer neutral; und das muss mit dem Gegenüber überhaupt nichts zu tun haben. Unsere Technologie ergänzt das durch eine neutrale Position. Fehlentscheidungen kann es so oder so geben, die lassen sich nie zu 100 Prozent ausschließen. Aber die Technologie kann als unabhängige Kontrollinstanz hinzukommen. Man muss diesen Prozess je nach Einsatzgebiet definieren.

Martin Schaarschmidt: Wie ist es mit dem Erkennen von Lügen?

Dagmar Schuller: Das ist sehr schwierig, denn eine Lüge ist an sich keine Emotion. Zu entscheiden, ob jemand lügt oder nicht, fällt auch uns Menschen viel schwerer als das Erkennen einer Emotion. An was will ich das Lügen fest machen, um es dann meinem System vermitteln zu können?

Martin Schaarschmidt: Aber Emotionen erkennt es. Wäre es dann auch denkbar, dass mir mein Hörgerät irgendwann meldet: „Pass auf, dein Gesprächspartner wird wütend!“ Oder: „Pass auf, dein Gesprächspartner schläft gleich ein!“?

Dagmar Schuller: Natürlich ist so etwas denkbar. Da geht es um prädiktive Analytik. Das wird ja vor allem im Gesundheitsbereich schon genutzt – etwa bei der Diagnose von Burn-out und Depressionen. Hier nutzen wir die Appraisal- bzw. Einschätzungs-Theorie von Professor Klaus R. Scherer, dem weltweit führenden Psychologen für Emotionserkennung. Es geht darum, zu verstehen, wie Emotionen verlaufen, wie sie sich auf- und wieder abbauen. Um möglichst präzise vorhersagen zu können, wie sich etwas auf emotionaler Ebene entwickelt. Marker, die z. B. entstehende Wut oder Ermüdung zeigen, sind da natürlich machbar. Und Hörgeräte haben den Vorteil, dass sie nicht nur einen Sensor bzw. ein Mikrofon haben, sondern mehrere.

Es gibt heute schon Oberflächen, mit denen Roboter wahrnehmen können. Wie mit einer Haut können sie Dinge erfühlen. Bei der Münchner Firma tacterion, einem Spin-Off des Deutschen Zentrums für Luft- und Raumfahrt (DLR), funktioniert das bereits sehr gut. Die statten Roboter mit solchen Oberflächen aus, deren Wahrnehmungsfähigkeit dann mit der von Fingerkuppen vergleichbar ist. Mit solchen dünnen Schichten könnte man auch Hörgeräte überziehen, um zusätzliche Aufnahmekomponenten zu schaffen. Man könnte die Temperatur am Ohr messen, den Pulsschlag, die Atemfrequenz. Je mehr Sensorik Sie anbringen können, umso mehr Outcome können sie haben – und damit das Ergebnis weiter verbessern.

Martin Schaarschmidt: Frau Schuller, haben Sie vielen Dank für das interessante Gespräch!

Weitere Informationen zu audEERING unter www.audeering.com.

Fotos: audEERING