

Digitalisierung & KI

Alexander Waibel



Leiter der Interactive System Labs am KIT FOTO: KIT

Alexander Waibel ist ein Pionier für maschinelle Übersetzungen. Bei einer Tauchfahrt zur Titanic führte er ein Experiment durch, das zeigen soll, wie man Sprachbarrieren in Videokonferenzen oder die Synchronisation von Filmen künftig abschaffen könnte.



von Oliver Voß

veröffentlicht am 03.08.2022

Mitte Juli hat sich **Alexander Waibel** einen besonderen Traum erfüllt. Der Professor und passionierte Taucher besichtigte an Bord einer Forschungsmission das Wrack der Titanic. Begeistert zeigt er das Foto eines der gigantischen Heizkessel, auf einem anderen sind die Kronleuchter im Ballsaal zu sehen. „Die hängen da immer noch“, sagt Waibel.

Ähnlich spektakulär ist ein kurzes Video vom Tauchgang. Es zeigt ein Gespräch mit **Stockton Rush**, dem Chef des Forschungs-U-Boot-Bauers Ocean Gate, über die Sauerstoff- und Stromversorgung. Allerdings ist das Video nicht echt, sondern ein **Deepfake aus 4.000 Metern Tiefe**.

Waibel hat aus dem U-Boot lediglich Textnachrichten an die Oberfläche geschickt. Dort werden sie von einem Computer per **Sprach- und Videosynthese** so umgewandelt, dass es auf dem Video so aussieht und klingt, als würde die Person selbst gerade sprechen - samt Lippenbewegungen. Dafür waren vor dem Tauchgang Videos und Stimmproben aufgezeichnet worden.

Waibel klingt tatsächlich sehr realistisch, sein Gesprächspartner hört sich hingegen noch mehr wie ein Computer als ein Mensch an. Es habe kaum Videomaterial zum Training von ihm gegeben, sagt Waibel entschuldigend. Und noch eine andere Problematik zeige das Experiment: „Es ist extrem schwierig, natürlich rüberzubringen, wenn eine Person nichts sagt.“ Damit das Video nicht wie eingefroren wirke, müssten auch leichte Gesichts- und Körperreaktionen erzeugt werden.

Bessere Kommunikation bei geringer Bandbreite

Größere Datenmengen vom Meeresgrund durch das Salzwasser zu übertragen ist bisher nicht möglich, daher wurde **Sonar** eingesetzt, doch das bietet nur eine **sehr niedrige Bandbreite**, die gerade für Textnachrichten reiche. Der Test soll daher zeigen, wie man trotz mangelnder Bandbreite die **Kommunikation verbessern** kann. „Das ist in vielen Bereichen ein Thema“, sagt Waibel. Angefangen von Extremsituationen wie in U-Booten oder der Luft- und Raumfahrt.

Doch auch im Alltag ist das Netz oft für viele Datenübertragungen zu schlecht. Auch dort könnten sich Anwendungsfelder ergeben. „Eventuell kann so eine Technologie auch eingesetzt werden, um **temporäre Aussetzer in Videokonferenzen zu überbrücken**“, sagt der Professor vom **Karlsruhe Institut of Technology (KIT)**.

Profitieren könnte dabei Zoom, denn der US-Konzern optimiert sein Angebot ohnehin schon mit dem Know-How aus Karlsruhe. So hat das

Unternehmen Ende Juni **Echtzeitübersetzungen** von Englisch in zehn Sprachen eingeführt, die Technologie dahinter stammt letztlich ebenfalls von Waibel und seinem Team.

Pionier bei maschinellen Übersetzungen

Der inzwischen 66-Jährige hatte 1990 als Professor in Karlsruhe begonnen und sich bereits zu dieser Zeit mit **Spracherkennung und maschinellen Übersetzungen** beschäftigt. Dabei setzte er schon damals auf den Einsatz von **neuronalen Netzen**. „Wir wurden damals oft als **neural nuts** beschimpft“, sagt Waibel.

Der Durchbruch des KI-Ansatzes kam dann erst mehr als zehn Jahre später mit der Verfügbarkeit größerer Rechenleistungen. Er war dann selbst über die Entwicklungssprünge verblüfft: Habe man vorher stolz Verbesserungen von 0,5 Prozent publiziert, erreichten die Wissenschaftler plötzlich 30 Prozent.

Das nutzte er beispielsweise für die **Entwicklung des ersten Echtzeit-Simultanübersetzungssystems für Vorlesungen**. Der „**Lecture Translator**“ startete 2013 am KIT. Auch an einem Übersetzungssystem für das EU-Parlament arbeitete er dann, 2018 erfolgte der Proof of Concept des **Projekts EU-Bridge** bei einer Konferenz. Doch mehr wurde nicht daraus.

Facebook und Zoom kauften seine Start-ups

Auch deswegen profitieren inzwischen primär US-Konzerne von den Erfindungen des Karlsruhers, der auch noch an der **Carnegie Mellon University (CMU)** in den USA arbeitet. Die von Waibel mitentwickelte **Übersetzungs-App Jibbigo** wurde 2013 von **Facebook** gekauft. Und im Vorjahr übernahm **Zoom** die von Waibel und Sebastian Stücker gegründete **Kites GmbH**. Das Unternehmen nutzt nun deren Know-How für die neue Übersetzungsfunktion und entwickelt solche Technologien auch am Standort Karlsruhe weiter, wo Zoom ein Forschungs- und Entwicklungszentrum aufbaut.

Für die Zukunft schwebt Waibel dabei jedoch deutlich mehr als simultan übersetzte Untertitel vor. Er verfolgt schon lange die **Vision einer**

„**sprachtransparenten Gesellschaft**“. Nun ist seiner Ansicht nach tatsächlich der Punkt erreicht, wo man mithilfe von Technologie die **Sprachbarrieren abschaffen** kann. So könnte es künftig Videokonferenzen geben, bei denen die Teilnehmer in ihrer Sprache sprechen und das Gegenüber hört eine Übersetzung – jedoch mit der echten Stimme und den Lippenbewegungen des Gesprächspartners. „Man merkt dann gar nicht mehr, dass die andere Person die Sprache gar nicht kann“, sagt Waibel. Genauso könnten womöglich Untertitel oder die **Synchronisierung in Filmen überflüssig** werden und stattdessen die Zuschauer die Originalstimme der Schauspieler in einer anderen Sprache hören.

Die entsprechende Technik ist nach Ansicht von Waibel schon relativ weit, wie auch der Titanic-Test gezeigt habe. Für gängige Sprachen sind solche Szenarien für den realen Einsatz daher nicht mehr weit: „Wenn man es forciert, könnte man das **in einem oder zwei Jahren** schaffen.“
Oliver Voß

Drei Fragen an Alexander Waibel:

Welche Innovation wünschen Sie sich?

Das Problem der Sprachbarrieren haben wir in meiner Forschergeneration relativ gut in den Griff bekommen. Allerdings nur für eine Handvoll Sprachen, doch insgesamt gibt es 7.000 in der Welt. Das für alle zu lösen ist ein hartnäckiges, offenes Problem, denn es ist eine Frage der Kosten und für viele Sprachen gibt es auch die Daten zum Training der Systeme nicht oder die Schriftsprache fehlt ganz. Das ist auch ein soziales Problem, weil wir so nur Lösungen für die studierte Elite entwickeln.

Wer aus der Digitalszene hat Sie beeindruckt?

Mark Raibert, der Chef von Boston Dynamics. Seine Bescheidenheit und stille Kompetenz hat mich immer beeindruckt. Und es war auch sehr mutig, seine Professur am MIT aufzugeben, um ein Start-up für humanoide Roboter zu gründen.

Als Digitalminister würde ich...

Abdanken. Aber eigentlich bräuchte Deutschland mehr Wissenschaftler in der Politik. Dort sind viele sogar auf das Gegenteil stolz. Doch es braucht mehr Fachexpertise und man muss mehr Visionen entwickeln, wie die DARPA, die verrückte Sachen machen kann. In Deutschland reden wir jetzt auch von **Sprunginnovationen**, doch das englische **Disruption beinhaltet Zerstörung**. Nur die Bereitschaft dazu ermöglicht Innovationen. Europa hätte dafür enorme Möglichkeiten, doch es fehlt der **Mut, auch mal ganze Industrien über den Haufen zu werfen**.