

Pressemitteilung

Institut für Informationssysteme forscht erfolgreich zur Sprachsynthese: **Menschliche Stimmen aus dem Computer – vom Original kaum noch unterscheidbar**

Hof, 29.03.2021 – Insbesondere für blinde oder sehbehinderte Menschen sind Computeranwendungen, die Texte vorlesen, bereits heute eine große Hilfe im Alltag. Auch beim Autofahren hat man sich längst an die freundlichen Stimmen aus der Navigation gewöhnt, die Fahrenden gefährliche Ablenkungen ersparen. Aber natürlich birgt die neue Technik auch Gefahren. Das Institut für Informationssysteme der Hochschule Hof forscht mit einer Studie an der Akzeptanz der künstlich generierten Stimmen und entwickelt eigene Modelle für den deutschen Markt.

Die Qualität der sogenannten Sprachsynthese hat sich in den letzten Jahren erheblich verbessert. Klängen die Stimmen lange Zeit eher blechern oder abgehackt, so weicht der Klang nach und nach einer zunehmenden Natürlichkeit und unauffälligen Sprachdynamik. Damit wird auch das Hören längerer Texte angenehm.

Rasante Verbesserung der Sprachqualität

„Erreicht werden konnte dies in der internationalen Forschung durch den Einsatz tiefer, neuronaler Netze. Insbesondere im englischen Sprachraum kann man bereits heute kaum noch unterscheiden, ob es sich um einen echten Menschen oder um ein Programm handelt“, so Prof. Dr. Rene Peinl, Leiter des Instituts für Informationssysteme der Hochschule Hof. Es gibt demnach mittlerweile eine Reihe frei verfügbarer Modelle, die sehr natürlich Englisch sprechen, wenn ausreichend viele Trainingsdaten verwendet werden. Die Spracherzeugung passiert dabei zumeist zweistufig. Zunächst wird ein sogenanntes Mel-Spektrogramm erzeugt, das eine Repräsentation der Sprachfrequenzen darstellt. Daraus erzeugt dann ein Vocoder das eigentliche Audiosignal. Beide Stufen sind neuronale Netze, die getrennt trainiert werden müssen.

Akzeptanz auf dem Prüfstand

Im DAMMIT-Programm der Hochschule Hof, das sich dem Technologietransfer zwischen Hochschule und mittelständischen Unternehmen zur digitalen Transformation auseinandersetzt, wird analysiert, wie hoch die Akzeptanz von Benutzern für computer-generierte Stimmen ist. Probanden bekommen dabei Textinhalte mittlerer Länge vorgelesen - also zum Beispiel Nachrichten mit einer halben Bildschirmseite Länge. Die in den letzten Jahren stattfindende, stetige Verbesserung der Qualität der Sprachsynthese erhöht einerseits den Komfort und die Einsatzmöglichkeiten der Technik, birgt andererseits aber auch Gefahren, da maschinelle Stimmen, die menschlich klingen, natürlich auch für Betrug oder Straftaten eingesetzt werden können.

Viele Anwendungsmöglichkeiten

Das automatisierte Vorlesen von Texten findet sich derzeit in immer mehr Anwendungsbereichen. Informationen aufnehmen zu können, obwohl sich die Augen auf ein anderes Ziel konzentrieren müssen, ist ein unschätzbare Vorteil: „Sprachsynthese ist natürlich ein wesentlicher Teil der Barrierefreiheit für Menschen mit Seheinschränkung. Ganz praktisch können aber zum Beispiel Aufträge für u.a. Gabelstaplerfahrer verbalisiert werden, was in deren Arbeitsablauf sehr hilfreich und zeitsparend sein kann. Oder man kann sich die täglichen Nachrichten mit der persönlichen Lieblingsstimme vorlesen lassen. Allgemein ist die Sprachsynthese auch ein wichtiger Teil von sprachgesteuerten Anwendungen wie Smart Speakern, z.B. Alexa von Amazon“, führt Prof. Dr. Peinl einige Einsatzmöglichkeiten aus.

Marktnachfrage wächst

Dabei dürfte die Nachfrage nach automatisch erstellten, aber menschlich klingenden Stimmen, erst am Anfang stehen. Ein Beispiel hierfür ist am Campus der Hochschule Hof und dort im Gründerzentrum Einstein 1 zu finden: Das Start-up-Unternehmen ahearo bietet einen Dienst an, bei dem man sich Inhalte, die sonst nur als Text verfügbar sind, auch als Audio-Podcast anhören kann. Diese Texte werden bisher von menschlichen Sprechern eingelesen. „Eine solche Produktion ist natürlich kostenintensiv und stößt auch aufgrund der eingeschränkten Verfügbarkeit von professionellen Sprechern an ihre Grenzen. Die Zusammenarbeit mit der Hochschule Hof eröffnet uns deshalb völlig neue Möglichkeiten“, so Johannes Garbarek, Gründer und CEO von ahearo.

Hohe Geschwindigkeit und geringe Kosten

„Für ahearo und andere Unternehmen, die eine kostengünstige und schnelle Möglichkeit suchen, qualitativ ansprechende Sprachsynthese in ihre Produkte einzubinden, entwickeln wir eine Lösung für die Generierung deutscher Sprache aus Text.“, so Prof. Dr. Peinl. Dabei kommen frei verfügbare, selbst erstellte und von ahearo bereitgestellte Audiodaten zum Einsatz, um die Sprachsynthese-Modelle bestmöglich zu trainieren. Die Evaluation erfolgt sowohl über objektiv messbare Werte als auch über subjektive Einschätzungen der Probanden.

Ermutigende Zwischenergebnisse

Die bislang vorliegenden Ergebnisse sind ermutigend und lassen auf einen baldigen praktischen Einsatz der Software hoffen: „Kurze Sätze werden in unserem Modell bereits sehr gut vorgelesen. Herausforderungen sind aktuell noch Pausen und Betonungen in komplexeren Sätzen, sowie Abkürzungen, zusammengesetzte Wörter sowie Eigennamen“, erläutert Forscher Peinl. Dass es dabei dem Computerprogramm mitunter nicht anders ergeht als dem Menschen, belegt eine kleine Anekdote: „Wir haben zum Beispiel das Wort „Frühsommer-Meningoenzephalitis (FSME)“ in unseren Test-Texten stehen. Wen wundert es, dass nicht nur wir, sondern auch der Computer mit solchen Wortungetümen so seine Schwierigkeiten hat“, so Professor Dr. Peinl.

Förderung

Die Ergebnisse der Studie, sowie die im Rahmen der Forschung entstehende Software werden frei veröffentlicht und auch zugänglich gemacht. Das Projekt wird im Rahmen des EFRE-Programms Bayern 2014-2020, von der Europäischen Union über den Fond für Regionale Entwicklung sowie vom Bayerischen Staatsministerium für Wissenschaft und Kunst gefördert. Ein weiterer Projektpartner ist zudem das am Campus der Hochschule Hof beheimatete Unternehmen smartlytic GmbH für Softwareentwicklung und Datenanalyse.

Pressekontakt:

Rainer Krauß, Hochschulkommunikation / PR
Alfons-Goppel-Platz 1, 95028 Hof
Telefon: 09281/409-3006
E-Mail: pressestelle@hof-university.de

Über die Hochschule Hof:

Praxisorientierung, Internationalisierung und intelligente Ressourcennutzung stehen im Fokus von Lehre und Forschung an der Hochschule Hof. Im Bereich Internationalisierung legt die Hochschule einen weiteren Schwerpunkt auf Indien, im Hinblick auf das Thema intelligente Ressourcennutzung stehen Wasser- und Energieeffizienz im Vordergrund. Das breitgefächerte und interdisziplinäre Studienangebot reicht von Wirtschaft und Wirtschaftsrecht bis hin zu Informatik und Ingenieurwissenschaften. Der Campus Münchberg bietet durch eng mit der Wirtschaft verzahnte Textil- und Designstudiengänge eine in Deutschland einmalige Ausbildung. Auch die hochfränkischen Unternehmen profitieren durch die Einrichtung von Kompetenzzentren und Instituten an der Hochschule. Die Schwerpunkte der vier Forschungsinstitute liegen auf den Bereichen Informationssysteme, Materialwissenschaften, Wasser- und Energiemanagement sowie Biopolymere. Am Institut für Weiterbildung finden berufstätige Fach- und Führungskräfte nationale als auch internationale Weiterbildungsangebote auf Hochschulniveau; das Programm des ifw beinhaltet dabei berufsbegleitende Bachelor- und Masterstudiengänge, Zertifikatslehrgänge, akademische Weiterbildungskurse und Seminare. Das an die Hochschule Hof angegliederte Bayerisch-Indische Zentrum für Wirtschaft und Hochschulen BayIND koordiniert und fördert die Zusammenarbeit zwischen Bayern und Indien. Studierende mit StartUp- oder Gründungsinteresse werden beraten und gefördert durch das Digitale Gründerzentrum Einstein1 am Campus der Hochschule.