



Alexander Waibel ist Professor an der KIT-Fakultät für Informatik sowie an der School of Computer Science an der Carnegie Mellon University, Pittsburgh, USA. Er ist Direktor von interACT, dem International Center for Advanced Communication Technologies. Das Netzwerk von neun führenden Forschungseinrichtungen weltweit, koordiniert gemeinsame Forschungs- und Austauschprogramme.

Die Forschungsinteressen von Professor Waibel konzentrieren sich auf Künstliche Intelligenz sowie multilinguale und multimodale Mensch-Maschine-Interaktionstechnologien. Sein Arbeiten war schon früh angetrieben von der Überzeugung, dass intelligente Systeme lernen müssen und dass komplexe Sprachtechnologien nur durch maschinelles Lernen und Interaktion möglich würden. Mit dem Time-Delay Neural Network (TDNN) stellte er bereits 1987 den ersten Ansatz vor, durch den Neuronale Netzwerke verschiebungsinvariant lernen können und lieferte damit einen der wichtigsten Grundbausteine heutiger künstlich intelligenter Systeme. Darauf aufbauend entwickelten er und sein Team neuartige Kommunikationstechnologien. Seit 2012 ist der automatische Echtzeit-Vorlesungsübersetzungsdienst als Service für ausländische Studierende am KIT im Einsatz.

Professor Waibel leitet und koordiniert zahlreiche internationale Forschungsprogramme. Neben seiner wissenschaftlichen Arbeit widmet er sich besonders dem Technologietransfer. Er ist Fellow der IEEE und Mitglied der Nationalen Akademie der Wissenschaften „Leopoldina“.

ÜBERBLICK UND ALLGEMEINES

Die Interactive Systems Labs (ISL) erforschen Technologien, welche die menschliche Kommunikation verbessern und Themen umfassen wie multimodale Schnittstellen, Spracherkennung, maschinelle Übersetzung, Sprachsynthese, Handschriftenerkennung, Mensch-Maschine-Interaktion, Neuronale Netze und Maschinelles Lernen.

Die Forschung führte zu innovativen Erstaufführungen und Durchbrüchen:

- TDNN, das erste „convolutional“ neural net (1987)
- der erste Echtzeit-Vorlesungsübersetzungsdienst der Welt (im Einsatz an Universitäten seit 2012)
- mobile Übersetzungssysteme für gesprochene und geschriebene Sprache auf Smartphones (Jibbiggo, 2009, Road Sign Translator, 2001)
- das erste Echtzeit-Simultanübersetzungssystem für Vorlesungen (Lecture Translator, 2005)
- erste Multimodale Benutzerschnittstellen: Lippenlesen (1993), Handschrifterkennung, Focus of Attention Tracking (1998), EMG Sprache (2005)

ERGEBNISSE UND ERFOLGE

Im Bereich der automatischen Spracherkennung entwickelt die Gruppe, geleitet von Dr. Sebastian Stüker, Spracherkennungssysteme für kontinuierliche Sprache mit großem Vokabular, insbesondere für den KIT Lecture Translator. Diese Systeme liefern kontinuierliche Ausgabe auf Streams von Sprachen, z.B. Vorträgen und Vorlesungen, mit einer sehr geringen Latenz. Um optimale Leistung für unterschiedliche Vorträge und Vorlesungen zu erzielen, liegt ein besonderer Fokus auf der Entwicklung von Adaptionstechniken von Sprachmodellen und Aussprachewörterbüchern, die auf heterogenen Daten operieren und möglichst wenig menschlicher Intervention bedürfen. Ferner entwickelt die Gruppe Techniken für die schnelle Portierung von Spracherkennungssystemen auf neue Sprachen mit wenig Trainingsdaten.

Im Bereich der maschinellen Übersetzung entwickelt die Gruppe unter der Leitung von Dr. Jan Niehues Übersetzungssysteme für Text und gesprochene Sprache. Motiviert durch den Erfolg von Deep-learning Methoden in der maschinellen Übersetzung, lag ein Fokus der Forschung in 2018 auf der Weiterentwicklung dieser Methoden und ihre Anpassung auf Probleme beim praktischen Einsatz. Ein Schwerpunkt war dabei die Entwicklung von neuronalen Modellen für die Sprachübersetzung mit minimaler Latenz.

Mittels „multitask-learning“ wurde ein Modell entwickelt, das bereits für Teilsätze Übersetzungen erzeugt.

Um die maschinelle Übersetzung für möglichst viele der weltweit mehreren tausend Sprachen verfügbar zu machen, wurde ein multi-linguales Übersetzungssystem entwickelt, das nicht mehr nur von einer Quellsprache in eine Zielsprache übersetzen, sondern zwischen vielen Sprachen übersetzen kann. Dies hat zum Vorteil, dass sich der personelle Aufwand der Entwicklung von Systemen für viele Sprachen verringert und weniger Trainingsdaten benötigt werden.

Im Jahr 2018 wurden folgende Projekte eingeworben:

- „European Live Translator“ (ELITR), EU-Programm Horizon 2020, Laufzeit 2019-2021: Automatische Untertitelung von Besprechungen und Konferenzen, maschinelle Übersetzung von Gesprochenem sowie automatisches Protokollieren.
- „RELATER“, BMBF-Projekt: Entwicklung eines tragbaren, sicheren und erweiterbaren Übersetzungssystem für diagnostische Interviews mit traumatisierten Flüchtlingen aus dem arabischen Sprachraum.
- Thematisches Netzwerk „CLICS“, DAAD-Projekt, Anschlussförderung 2019-2020: Massive Open Online Courses (MOOC) in die Lehre bringen, Austausch, Vorträge.

AUSGEWÄHLTE PUBLIKATIONEN

M. Müller, S. Stüker, A. Waibel.: Neural Language Codes for Multilingual Acoustic Models. Interspeech 2018.

T.-L. Ha., J. Niehues, A. Waibel.: Toward Multilingual Neural Machine Translation with Universal Encoder and Decoder. In: *Proceedings of the 13th International Workshop on Spoken Language Translation (IWSLT)*. Seattle, USA, 2016.

A. Waibel, R. Stiefelhagen (Eds.): Computers in the Human Interaction Loop. Springer, 2009.

C. Fügen, A. Waibel, M. Kolss: Simultaneous translation of lectures and speeches. In: *Journal of Machine Translation*. Vol. 21, No. 4, Springer, Niederlande, S. 209-252, 2008.

A. Waibel, T. Hanazawa, G., Hinton, K. Shikano, K. Lang: Phoneme Recognition, Using Time-Delay Neural Networks. IEEE Transactions of the Acoustics, Speech and Signals Processing Society. Vol. 37, No. 3. 1989.

MITARBEITERINNEN UND MITARBEITER

Verwaltungspersonal

Silke Dannenmaier (Sekretariat)
Margit Rödder (Öffentlichkeitsarbeit)

Wissenschaftliches Personal

Stefan Constantin
Thanh-Le Ha
Markus Müller
Thai Son Nguyen
Jan Niehues
Teven Le Scao
Ngoc Quan Pham
Elizabeth Salesky
Felix Scheider
Ilya Shibayev
Matthias Sperber
Sebastian Stüker
Thomas Zenkel
Zhong Zhou