



[Prof. Dr. Alex Waibel // Interactive Systems Labs]

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

Die Forschungsinteressen von Alex 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 nur durch maschinelles Lernen und Interaktion möglich würden. Mit dem Time-Delay Neural Network (TDNN) stellte er 1987 das erste verschiebungsinvariante („convolutional“) Neuronale Netz vor, einer der wichtigsten Grundbausteine heutiger KI. Darauf aufbauend entwickelten er und sein Team mehrere neue Technologien zur Überwindung von Sprach-Barrieren: u. a. erster Dialog-Sprachübersetzer als Prototyp (1991) und Produkt (2009), erster Simultan-Übersetzer von Vorlesungen als Prototyp (2005), als Dienst für Studierende (2012) und als Produkt (2015). Neben seiner wissenschaftlichen Arbeit widmet er sich besonders dem Transfer neuer Technologien in die Industrie, öffentliche und humanitäre Dienste. Er war Gründer von über 10 Startups, die von Firmen (Facebook, Zoom, 3M, ...) übernommen wurden, oder zu einem Börsengang führten.

Waibel ist Fellow der IEEE und ISCA, sowie Mitglied der Nationalen Akademie der Wissenschaften „Leopoldina“. Seine Arbeiten sind in über 900 Publikationen (h-index 94) dokumentiert und wurden durch die 2023 IEEE Flanagan Field Medal, sowie zahlreiche weitere Preise gewürdigt.

// Überblick und Allgemeines

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

Das übergreifende Ziel der Forschungsarbeiten ist das Überwinden von Kommunikationsbarrieren zwischen Menschen durch lernende KI-Systeme, die simultan und konsekutiv Dolmetschleistungen in unterschiedlichen Situationen und Lebenslagen übernehmen können. Der „Lecture Translator“ ist hierzu die Realisierung langjähriger Vorarbeiten: Gesprochene Sprache simultan, mit mensch-ähnlicher Genauigkeit, schritthaltend bei niedrigster Latenz in beliebigen Sprachen erkennen und übersetzen zu können; und dies in die Zusammenarbeit zwischen Menschen so einzuführen, dass es zunehmend zu einem sprachensprache- bzw. barrierefreien Miteinander führt. (Abb. 1) Die zugrunde liegenden Algorithmen und Systeme unterstützen ausländische Studierende im Vorlesungsbetrieb am KIT und an Partneruniversitäten, sie fertigen simultane Untertitelung für Hörbehinderte an und vermitteln bei Arzt-Patienten-Dialogen mit Migrantin-

1
Kommunikation über verschiedene Sprachen und Barrieren hinweg – möglich durch die Technologien entwickelt in den Interactive Systems Labs



nen und Migranten (wie z. B. in Zusammenarbeit mit dem Zentralinstitut für Seelische Gesundheit in Mannheim). Während des COVID-19-Lockdowns 2020 konnten die Simultanübersetzer am KIT sogleich ohne Umstellung ZOOM-Lehre und Video-Konferenzen in verschiedenen Sprachen durch automatische Transkription, Übersetzung und Suchbarkeit sowohl in Echtzeit als auch in archivarischem Material beim Studium anbieten.

// **Forschung und Erfolge**

Durch das automatische neuronale Lernen in der Sprache konnte die nötige Qualität, Geschwindigkeit und Skalierbarkeit der Sprach-Erkennung (nun erstmalig besser als der Mensch!), -Verarbeitung, und -Übersetzung erreicht werden, um das nötige Sprachwissen automatisch/datengetrieben aufzunehmen, Anpassungsfähigkeit, und multilinguale Skalierbarkeit auf vielen Sprachen und Integration in menschen-freundlichen Benutzerschnittstellen sicher zu stellen. Dies konnte jüngst erreicht werden, so dass die Systeme schon nach kurzer Zeit in Deutschland, Europa und weltweit erfolgreich zum Einsatz kommen, um ein globales Miteinander zu vereinfachen und zu ermöglichen. Auch die Umsetzung dieser Grundlagen ist auf Wirken des ISL zurückzuführen, da Ausgründungen aus dem Institut in Deutschland, USA und Japan nun den Kern von Sprachsystemen bei Facebook, ZOOM, 3M und anderen Unternehmen bilden.

// **Ausgewählte Publikationen**

“Super-Human Performance in Online Low-latency Recognition of Conversational Speech”, T.-S. Nguyen, S. Stueker, A. Waibel, INTERSPEECH 2021, Brno, Czech Republic, Aug. 30–Sep. 03, 2021

“Very Deep Self-Attention Networks for End-to-End Speech Recognition”, N.-Q. Pham, T.-S. Nguyen, J. Niehues, M. Müller, S. Stüker, A. Waibel, The 20th Annual Conf. of the Int. Speech Communication Association, INTERSPEECH, Austria, 2019.

“Multimodal Dialogue Processing for Machine Translation”, A. Waibel, in: Handbook for Multimodal Dialogue, S. Oviatt, Ph. Cohen, eds., Morgan & Claypool (2019), pgs. 557 – 620.

“Low-Latency Neural Speech Translation”, J. Niehues, N.-Q. Pham, T.-L. Ha, M. Sperber, A. Waibel, INTERSPEECH 2018, Hyderabad, India, Sept. 2–6, 2018.

“Lecture Translator – Speech Translation Framework for Simultaneous Lecture Translation”, M. Müller, T.-S. Nguyen, J. Niehues, E. Cho, B. Krüger, T.-L. Ha, K. Kilgour, M. Sperber, M. Mediani, S. Stüker, A. Waibel, Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), USA, 2016.

// **Mitarbeiterinnen und Mitarbeiter**

Verwaltungspersonal

Silke Dannenmaier (Sekretariat)
 Margit Rödder (Presse und Öffentlichkeitsarbeit)

Wissenschaftliches Personal

Leonard Bärmann
 Stefan Constantin
 Hazim Ekenel
 Christian Huber
 Aya Khader
 Mohamed Mediani
 Carlos Mullov
 Thai Binh Nguyen
 Tuan-Nam Nguyen
 Ngoc Quan Pham
 Fabian Retkowski
 Enes Yavuz Ugan
 Dogucan Yaman
 Fevziye İrem Yaman

// **Website**

isl.anthropomatik.kit.edu