

Vortrag von Herrn M.Sc. Stefan Taubert Professur Medieninformatik

„Vergleich des Einflusses impliziten und expliziten Lernens der Lautdauer auf die Verständlichkeit und Natürlichkeit von Sprachsynthese in neuronalen Systemen“

Sprachsynthese (Text-to-Speech, TTS) bezeichnet die computergestützte Umwandlung von Text in Sprache, wobei Verständlichkeit und Natürlichkeit zentrale Qualitätskriterien sind. Neuronale Ansätze haben die Sprachqualität in den letzten Jahren deutlich verbessert, ermöglichen jedoch bislang keine präzise Kontrolle der Lautdauer auf Phonemebene.

Diese Arbeit untersucht, ob ein neuronales TTS-System bereits während des Trainings explizit lernen sollte, die Lautdauer auf Phonemebene zu steuern, oder ob ein rein implizites Lernen aus umfangreichen Sprachdaten ausreicht. Hierfür wurde ein bestehendes TTS-System erweitert, um mithilfe automatischer Lautdauerannotation explizit zu lernen. Dazu erfolgten eine phonemische Transkription der Aufnahmen, die zeitliche Ausrichtung sowie die Einteilung der Lautdauern in vier Quantile.

Eine subjektive Evaluation über Amazon Mechanical Turk zeigte, dass der explizite Ansatz für Englisch die Natürlichkeit bei gleichbleibender Verständlichkeit hochsignifikant verbesserte. Objektive Messungen stützten die qualitative Überlegenheit des expliziten Ansatzes. Eine ergänzende Nebenstudie belegte zudem, dass mit beiden Ansätzen auch für Mandarin eine natürliche und verständliche Sprache erzeugt werden konnte.

Die Methode erhöht die Steuerbarkeit und Natürlichkeit neuronaler Sprachsynthese und eröffnet neue Möglichkeiten für die linguistische Forschung.

Öffentliche Verteidigung im Rahmen des Promotionsverfahrens

“Vergleich des Einflusses impliziten und expliziten Lernens der Lautdauer auf die Verständlichkeit und Natürlichkeit von Sprachsynthese in neuronalen Systemen“

18. September 2025, 09:00 Uhr

Technische Universität Chemnitz, Straße der Nationen 62, Raum: A10.367 (alt: 1/367)

Alle interessierten Personen sind eingeladen.