STEFAN TAUBERT

Research Engineer – PhD in Text-to-Speech (TTS)

stefantaubert · stefan-taubert · contact@stefantaubert.com · 0000-0002-4932-2874
Chemnitz, Saxony, Germany

EXPERIENCE

Research Scientist

Chemnitz University of Technology

Mar 2020 – Present — Chemnitz, Germany

- Researcher in the German Research Foundation (DFG) Collaborative Research Center (CRC) "Hybrid Societies," subproject "Credibility Through Non-native Language Varieties in Conversational Pedagogical Agents"; improving synthetic-voice credibility by modeling and integrating grammatical, segmental, and temporal language variety cues into TTS.
- Built an end-to-end pipeline for non-native language varieties: recording, annotation, analysis, training, and synthesis of custom TTS voices (speakers from China, Italy, Nigeria, Turkey, and the Czech Republic).
- Wrote PhD thesis on explicit phoneme-duration control in neural TTS; controlled listening study (n = 33) showed a significant naturalness improvement vs. baseline (p = 0.0048); MOS ratings: intelligibility 4.44 (ground truth 4.63), naturalness 3.55 (ground truth 4.17).

Software Developer

IPML Institut für Produktionssteuerung, Materialwirtschaft und Logistik GmbH

Jun 2013 – Feb 2020 (part-time) — Chemnitz, Germany

- Developed software components for discrete-event simulation and ERP related software in C# and .NET.
- Authored developer/user documentation, administered Confluence, and implemented software tests.

Research Assistant

Chemnitz University of Technology

Mar–Aug 2018 & Aug–Sep 2019 (part-time) — Chemnitz, Germany

- Developed methods to identify plant species from images and environmental variables using ML (LifeCLEF 2018, 2nd place).
- Built a method for automated lifelog moment retrieval from images and sensor data using ML (ImageCLEF 2019).

EDUCATION

PhD in Text-to-Speech

Chemnitz University of Technology

Mar 2020 – Oct 2025

Thesis topic: Comparison of implicit and explicit phoneme duration learning and its impact on intelligibility and naturalness of neural speech synthesis.

BSc & MSc in Applied Computer Science

Chemnitz University of Technology

Oct 2014 – Sep 2019

- Bachelor thesis: Duplicate question detection in Q&A platforms using machine learning (PDF).
- Master thesis: Automated lifelog moment retrieval based on image segmentation and similarity measures (PDF).

AWARDS

Deutschlandstipendium Scholarship
Award for academic excellence.

Best Paper Award – 2nd Place
Novel vowel quality evaluation method for TTS (IEEE CIVEMSA 2022).

GITHUB TOP REPOS

mel-cepstral-distance

☆ 57 · 10 · Python

Calculate MCD between two inputs for objective evaluation of TTS quality.

pinyin-to-ipa

☆ 51 · 10 · Python

Transcribe pinyin (Chinese phonetic alphabet) into IPA for TTS applications.

200+ Stars on GitHub
20+ PyPI packages with 25+ forks

LANGUAGES

German
Native

English
C1 (business fluent)

TEAM & LEADERSHIP

- Supervised three research assistants.
- Taught two university courses (media applications and media retrieval), one semester each with 10–20 students.
- Collaborated in an interdisciplinary team (linguistics, psychology) within the CRC.
- Supervised a master's thesis on speaker identification using AI.

SKILLS

Analytical thinking · Eye for detail · Holistic problem-solving · Teamwork

Python · PyTorch · NumPy · NLP · Data analysis · C#

PUBLICATIONS

Book Chapters

- Albrecht, S., Tamboli, R., **Taubert**, **S.**, Meusel, F., Eibl, M., Rey, G. D., & Schmied, J. (2025, in press). Establishing conversational pedagogical agents as credible knowledge providers: The case of synthesized Italian English. In B. Meyer, U. Thomas, & O. Kanoun (Eds.), *Hybrid Societies: Humans interacting with embodied technologies* (Vol. 1). Chemnitz, Germany: Springer.
- Eibl, M., Haupt, J., Kahl, S., **Taubert**, **S.**, & Wilhelm-Stein, T. (2022). C 6 Audio- und Musik-Retrieval. In *Grundlagen der Informationswissenschaft* (pp. 423–432). doi:10.1515/9783110769043-036

Conference Proceedings

- **Taubert**, **S.**, & Eibl, M. (2025). Herausforderungen des Crowdworkings zu Forschungszwecken: Spam-Erkennung und Validierung in Amazon-Mechanical-Turk-Studien. In *18. Internationales Symposium für Informationswissenschaft (ISI 2025)*. doi:10.5281/zenodo.14925656
- Albrecht, S., Tamboli, R., **Taubert**, **S.**, Eibl, M., Rey, G. D., & Schmied, J. (2022). Towards a vowel formant based quality metric for text-to-speech systems: Measuring monophthong naturalness. In *2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* (pp. 1–6). doi:10.1109/CIVEMSA53371.2022.9853712
- Meyer, S., Albrecht, S., Eibl, M., Rey, G. D., Schmied, J., Tamboli, R., **Taubert**, **S.**, & Gesmann-Nuissl, D. (2022). Untying the Gordian knot: Legally compliant sound data collection and processing for TTS systems in China. In S. Schiffner, S. Ziegler, & A. Quesada Rodriguez (Eds.), *Privacy Symposium 2022* (pp. 87–104). doi:10.1007/978-3-031-09901-4_6
- **Taubert**, **S.** (2022j). Synthese von akzentspezifischen Sprachmerkmalen mittels phonetischer Transkriptionen. In D. Kowerko, M. Ritter, T. Köhler, & K. Helle (Eds.), *Proceedings of the 4th Interdisciplinary PhD Workshop of Media Informatics & Social Sciences 2022* (pp. 1–14). Laubusch, Germany: TUDpress.
- **Taubert**, **S.**, Sternkopf, J., Kahl, S., & Eibl, M. (2022). A comparison of text selection algorithms for sequence-to-sequence neural TTS. In *2022 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* (pp. 1–6). doi:10.1109/ICSPCC55723.2022.9984283
- **Taubert**, **S.**, Kahl, S., Kowerko, D., & Eibl, M. (2019). Automated lifelog moment retrieval based on image segmentation and similarity scores. In *CLEF 2019 Working Notes* (Vol. 2380, pp. 1–17). Università della Svizzera Italiana (USI), Switzerland. Retrieved from http://ceur-ws.org/Vol-2380/paper_83.pdf
- **Taubert**, **S.**, Mauermann, M., Kahl, S., Kowerko, D., & Eibl, M. (2018). Species prediction based on environmental variables using machine learning techniques. In *CLEF 2018 Working Notes* (Vol. 2125, pp. 1–10). Avignon, France. Retrieved from http://ceur-ws.org/Vol-2125/paper_93.pdf

TTS Models

- **Taubert**, **S.** (2023c). *LJ Speech English TTS*. Zenodo, 2023. doi:10.5281/zenodo.10200954
- **Taubert**, **S.** (2023d). *LJ Speech English TTS with explicit duration markers*. Zenodo, 2023. doi:10.5281/zenodo.10107103
- **Taubert**, **S.** (2023i). *THCHS-30 Chinese TTS*. Zenodo, 2023. doi:10.5281/zenodo.10210309
- **Taubert**, **S.** (2023j). *THCHS-30 Chinese TTS with explicit duration markers*. Zenodo, 2023. doi:10.5281/zenodo.10209989

Datasets

- **Taubert**, **S.** (2025). *LJS-MOS-120 [Audio]*. Hugging Face, 2025. doi:10.57967/hf/5368

Transcriptions

- **Taubert**, **S.** (2022f). *LJ Speech - aligned IPA transcriptions*. Zenodo, 2022. doi:10.5281/zenodo.7356907
- **Taubert**, **S.** (2023h). *THCHS-30 - aligned IPA transcriptions*. Zenodo, 2023. doi:10.5281/zenodo.7528595
- **Taubert**, **S.** (2023a). *CMU pronouncing dictionary with IPA transcriptions*. Zenodo, 2023. doi:10.5281/zenodo.7500804
- **Taubert**, **S.** (2023f). Pinyin - IPA mapping. doi:10.5281/zenodo.7525638

Software[1]

- **Taubert**, **S.** (2024a). En-tts [Computer software]. doi:10.5281/zenodo.10479347
- **Taubert**, **S.** (2024b). Zho-tts [Computer software]. doi:10.5281/zenodo.10512789
- **Taubert**, **S.** (2023e). Mean-opinion-score [Computer software]. doi:10.5281/zenodo.7670648
- **Taubert**, **S.** (2022k). Tacotron-cli [Computer software]. doi:10.5281/zenodo.7044284

[1] Just a selection; full list available at PyPI (pypi.org/user/stefantaubert)