Advancing Text-to-Speech Synthesis in Code-Switching Scenarios (Bachelor Thesis)

In the realm of human-robot interaction and multimedia content creation, text-to-speech (TTS) models play a pivotal role in crafting immersive and engaging experiences. Whether integrated into robotic systems or utilized for tasks like Face-dubbing and video generation, the quality of TTS significantly influences the perceived realism and overall effectiveness of the synthetic content.

This research endeavor seeks to delve into the landscape of state-of-the-art (SOTA) approaches [7, 10, 6, 2], examining their efficacy in the domain of Code-Switching (CSW) speech synthesis. Our primary objectives encompass:

- **Dataset Preparation:**
  - Investigate diverse SOTA models employing varied input-output configurations.
  - Implement data augmentation techniques to enhance dataset robustness.

- **Model Training:**
  - Implement and train models tailored for speech synthesis.

- **Performance Evaluation:**
  - Conduct quantitative assessments to gauge the models’ performance.
  - Evaluate and identify the most promising model for this specific application.

Throughout the research journey, you will receive guidance from experts specializing in Code-Switching speech recognition and multilingual speech synthesis. Access to cutting-edge GPUs will provide a robust foundation for your experiments.

Successful outcomes of this study will be compiled into a research paper for submission to a prestigious conference. If you are intrigued by this research opportunity, please submit your application, including your CV and transcript of records, to:

Enes Ugan
Email: enes.ugan@kit.edu
Website: [https://isl.anthropomatik.kit.edu/english/21_9532.php](https://isl.anthropomatik.kit.edu/english/21_9532.php)
Related research

3 Improving Code-Switching and Named Entity Recognition in ASR with Speech Editing based Data Augmentation
4 Zero-shot code-switching ASR and TTS with multilingual machine speech chain

References


