|  |  |  |  |
| --- | --- | --- | --- |
| **Datasets** | **Characteristics** | **Characteristics**  **and models** |  |
| CVSS | **CVSS** is a multilingual-to-English speech to speech translation corpus, covering sentence-level parallel S2ST pairs from 21 languages into English. CVSS is derived from the [Common Voice](https://paperswithcode.com/dataset/common-voice) speech corpus and the [CoVoST](https://paperswithcode.com/dataset/covost) 2 speech-to-text translation (ST) corpus, by synthesizing the translation text from CoVoST 2 into speech. | PnG NAT  PnG NAT with VC  Speaker Encoder |  |
| MuST-C | **MuST-C** currently represents the largest publicly available multilingual corpus (one-to-many) for speech translation. It covers eight language directions(English to German, Spanish, French, Italian, Dutch, Portuguese, Romanian and Russian). The corpus consists of audio, transcriptions and translations of English TED talks, and it comes with a predefined training, validation and test split. | analysis of adapters for multi-  lingual speech translation (ST). Starting from  different pre-trained models trained on non-parallel multilingual  data), we show that adapters can be used to: (a)  efficiently specialize ST to specific  language  Bilingual vs. Multilingual  Adapter tuning vs. Fine-tuning  Low-resource scenario  Fine tuning |  |
| MaSS | MaSS (Multilingual corpus of Sentence-aligned Spoken utterances) is an extension of the CMU Wilderness Multilingual Speech Dataset, a speech dataset based on recorded readings.  MaSS extends it by providing a large and clean dataset of 8,130 parallel spoken utterances across 8 languages. (The covered languages are: Basque, English, Finnish, French, Hungarian, Romanian, Russian and Spanish) | SpecAugment, a simple data augmentation method  for speech recognition is applied directly to  the feature inputs of a neural network.  LAS Network Architectures  Learning Rate Schedules  Shallow Fusion with Language Models |  |