1. Overview

**Motivation**
- Quality of pronunciation dictionary is important for Speech Recognition.
- Dictionaries may be of different quality depending on the creation process.

**Goal**
- Propose completely automatic methods to detect, remove, and substitute inconsistent and flawed dictionary entries.

**Data**
1. *Wiktionary* word-pronunciation pairs (provided by Internet community) that are used to build grapheme-to-phoneme (g2p) models.
   - Languages: English (en), German (de), Polish (pl), Spanish (es), Czech (cs), French (fr)
2. *GlobalPhone* Hausa (ha-GP) pronunciation dictionary (created by native speakers)
3. Singaporean English pronunciations (that have been generated with rules) to complement our Mandarin-English *SEAME* code-switch dictionary.

2. Automatic rejection of inconsistent and flawed entries

- Our investigated methods to filter erroneous entries fall into the following categories:

  1. **Length Filtering (Len)**
     Remove a pronunciation if the ratio of grapheme and phoneme tokens exceeds a certain threshold.

  2. **Epsilon Filtering (Eps)**
     a. Perform a 1-1 g2p alignment (Martirosian and Davel, 2007) (Black et al., 1998) which involves the insertion of graphemic and phonemic nulls (epsilons) into the lexical entries of words.
     b. Remove a pronunciation if the proportion of graphemic and phonemic nulls exceeds a threshold.

  3. **m-n Alignment Filtering (M2NALign)**
     a. Perform an m-n g2p alignment (Martirosian and Davel, 2007) (Black et al., 1998).
     b. Remove a pronunciation if the alignment score exceeds a threshold.

  4. **g2p Filtering (G2P)**
     a. Train g2p models with “reliable” word-pronunciation pairs.
     b. Apply the g2p models to convert a grapheme string into a most likely phoneme string.
     c. Remove a pronunciation if the edit distance between a synthesized phoneme string and the pronunciation in question exceeds a threshold.

- The threshold for each filtering method depends on:
  - the mean (μ) and
  - the standard deviation (σ) of the measure in focus.

- Those word-pronunciation pairs whose resulting number is shorter than μ - σ or longer than μ + σ are rejected.

- We built new g2p models with the remaining word-pronunciation pairs and applied them to the words with rejected pronunciations.

3. Results

- WER reduction of 27.3% relative on en.wikt word-pronunciation pairs with M2NALign.
- Without this outlier, the avg. improvement in wikt is 2.5% relative.
- On ha-GP, WER reduction with all filtered dictionaries but G2P by 1.5% relative on average. G2P len performs best with 2.6% relative improvement.
- On SEAME, we are able to slightly reduce the mixed error rate by 0.2% relative on average with a decoding using the filtered new pronunciations.

---

Automatic Error Recovery for Pronunciation Dictionaries
Tim Schlippe, Sebastian Ochs, Ngoc Thang Vu, Tanja Schultz
tim.schlippe@kit.edu