Self-Supervised Learning (SSL) pipeline for dysarthric speech proposal

Assume we have access to dysarthric speech audio applicable to our use case (i.e. **in-domain data**).

Setup and tune Voice Activity Detection (VAD). Adjust the **trimming of ‘long silences’ to higher than 1s**, as dysarthric speech has higher pauses, and **adjust a lower dB threshold** for what is considered noise since speech may be slurred.

Setup and train Audio Event Detection (AED) that *minimally* detects the difference between **`clean speech` events** vs **the rest** (even speech with background noise). The paper’s AED can classify a lot of audio events, but its experiment results show that we just need to focus on `clean speech` vs `non-clean speech` or `no speech`. Ensure the dataset trained on is updated such that **mumbling/slurring events will be classified as clean speech**, before training.

Follow preprocessing steps in Section 3.1, but set a **higher processing window shift** and **window shift**, say 40ms and 20ms respectively, since dysarthric speech may take a longer time. Apply **speech-crop for `clean speech`** and also **rand-crop**.

Follow **Lfb2vec procedure** to train a **multi-headed** model with **FlatNCE** objective. Have one Language Model (LM) head for each language targeted (e.g. English, Chinese, Malay, Tamil in the Singapore context).

Continuous Learning

Set up a **human-in-the-loop** workflow. One way is to create a UI that facilitates collecting both **explicit feedback** (👍👎) and **implicit signals**. Use **Active Learning** techniques to help **prioritize novel/unique data points to label**, and **invite users to participate/contribute** to label them.

It is important to ensure/**remind annotators** to **label with the proper intended sentences**, instead of the phonetics (e.g. `that` instead of `dat`, or `’cos` instead of `’cooooos` when one drags).

During training, use **regularization techniques** like EWC (Elastic Weight Consolidation) and **periodically fine-tune** on recent data (e.g. moving window) to combat catastrophic forgetting.