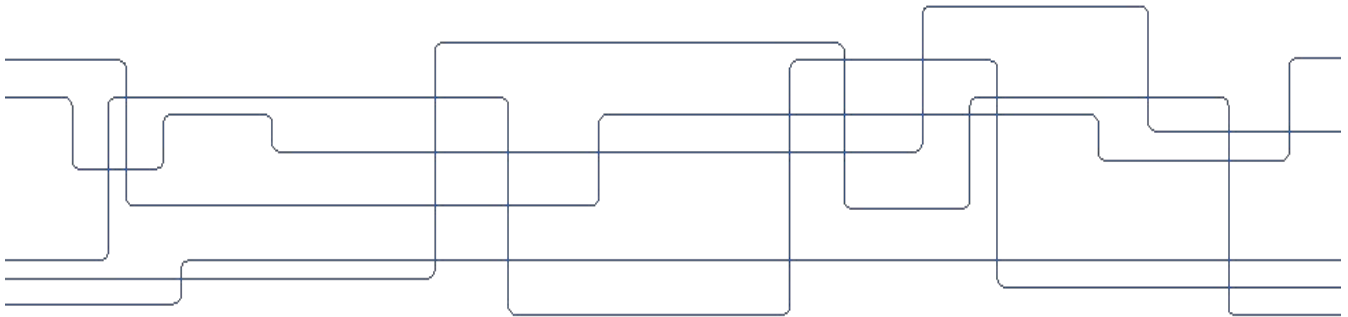# ASR fine tuning for minority languages and speaker adaptation

Jim O'Regan

Trend

Pretrained base model

(Usually using self-supervised learning on large amounts of data)

Fine-tuned for a particular task

(Usually using a much smaller amount of task-specific data)

Example: wav2vec2

Base model trained on large amounts of audio

Multilingual models:

XLSR-53

- Trained on 53 languages
- Generalises to others

XLS-R:

- Trained on 128 languages
- Generalises, but less well

    (Higher proportion of English training data)

Minority languages of Sweden

- Finnish
- Sami
- Romani
- Yiddish
- Meänkieli

Minority languages of Sweden

- ~~Finnish~~
- Sami
- *Romani*
- *Yiddish*
- Meänkieli

Finnish is already well taken care of

Romani and Yiddish lack sufficient (Swedish) data; data from other places where they are spoken may be sufficient

Sami and Meänkieli have representation in Swedish radio broadcasts.

Aim:

Provide speech infrastructure for (at least some) minority languages

Strategy: Meänkieli

Finnish ASR, plus regular orthographic transformations

(Meänkieli diverged from Finnish, but retains a high degree of mutual intelligibility)

Strategy: Sami

Phonetic triangulation (the traditional approach)

Use (weighted) phonetic models of multiple languages to find a likely phonetic transcript

Speaker Adaptation

Speaker Adaptation
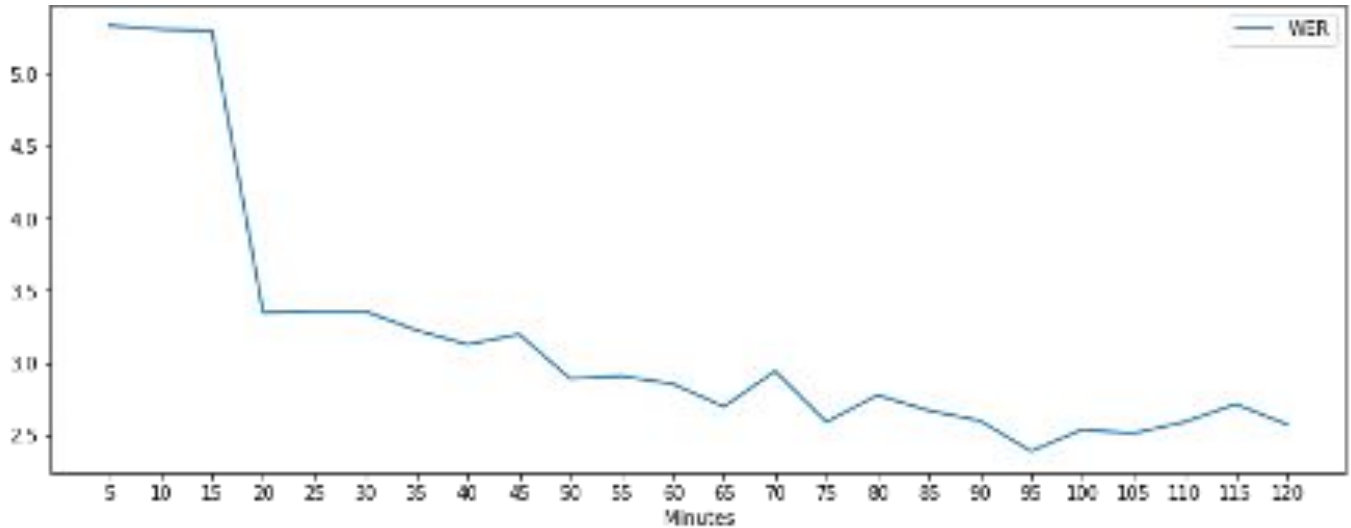
Improves speech recognition for individuals

Speaker Adaptation: Examples

Riksdag (Parliament) recordings

Folklore recordings

People of historical interest

Continuing for 300 updates

300 vs 1000 updates