A Comparison of Neural Network Methods for Unsupervised Representation Learning

Daniel Renshaw*, Herman Kamper*, Aren Jansen^, Sharon Goldwater*
(*University of Edinburgh, ^Johns Hopkins University)

1. Train stacked autoencoder
2. Unsupervised term discovery
3. Align word pair frames (DTW)
4. Train correspondence autoencoder

Input features

Unsupervised feature extractor
Track 1: unsupervised subword modeling
Xitsonga, across-speaker, ABX error rates
MFCC baseline: 33.8
MFCC UTD + cAE: 19.3 (43% reduction)
Supervised DNN, with bottleneck, trained on English, applied to Xitsonga without adaptation
Bottleneck features UTD + cAE: 18.2 (46% reduction)