

Perceptual compensation for tonal context in self-supervised speech models

Michele Gubian, Ioana Krehan, James Kirby

IPS, LMU München

This study examines the extent to which the wav2vec2.0 architecture exhibits evidence of compensation for phonological context. We conducted a pseudo-replication of a perceptual compensation experiment on Mandarin Chinese tones, and compared the embedding similarities and probing classifier outputs between a purely self-supervised pre-trained model and a model fine-tuned for Mandarin ASR. No evidence of compensation was found in the embedding similarities of the purely pre-trained model. Probing classifiers showed some evidence of compensation in addition to the expected layer-wise improvements in categorization, but failed to replicate human performance on isolated test syllables. Our findings contrast with previous reports of sensitivity to phonological structure emerging through pre-training alone, and suggest that supervised objectives may be necessary to encourage the abstraction of at least some types of phonological regularities.