

4aSCb28. Normalized recognition of speech and audio events

Session: Thursday Morning, Nov 03

Author: Mark A. Hasegawa-Johnson

Location: ECE Dept., Univ. of Illinois. Urbana, IL 61801

Author: Jui-Ting Huang

Location: ECE Dept., Univ. of Illinois. Urbana, IL 61801

Author: Sarah King

Location: ECE Dept., Univ. of Illinois. Urbana, IL 61801

Author: Xi Zhou

Location: ECE Dept., Univ. of Illinois. Urbana, IL 61801

Abstract:

An invariant feature is a nonlinear projection whose output shows less intra-class variability than its input. In machine learning, invariant features may be given a priori, on the basis of scientific knowledge, or they may be learned using feature selection algorithms. In the task of acoustic feature extraction for automatic speech recognition, for example, a candidate for a priori invariance is provided by the theory of phonological distinctive features, which specifies that any given distinctive feature should correspond to a fixed acoustic correlate (a fixed classification boundary between positive and negative examples), regardless of context. A learned invariance might, instead, project each phoneme into a high-dimensional Gaussian mixture supervector space, and in the high-dimensional space, learn an inter-phoneme distance metric that minimizes the distances among examples of any given phoneme. Results are available for both tasks, but it is not easy to compare them: learned invariance outperforms a priori invariance for some task definitions, and underperforms for other task definitions. As future work, we propose that the a priori invariance might be used to regularize a learned invariance projection. 12 1822