This paper explores the generality of COgnitive Component Analysis (COCA), which is defined as the process of unsupervised grouping of data such that the ensuing group structure is well-aligned with that resulting from human cognitive activity. The hypothesis of [COCA] is ecological: the essentially independent features in a context defined ensemble can be efficiently coded using a sparse independent component representation. Our devised protocol aims at comparing the performance of supervised learning (invoking cognitive activity) and unsupervised learning (statistical regularities) based on similar representations, and the only difference lies in the human inferred labels. Inspired by the previous research on COCA, we introduce a new pair of models, which directly employ the independent hypothesis. Statistical regularities are revealed at multiple time scales on phoneme, gender, age and speaker identity derived from speech signals. We indeed find that the supervised and unsupervised learning provide similar representations measured by the classification similarity at different levels.