Acoustic Representations Support Statistical Learning of Syllable Sequences: A Computational Study

Nika Jurov (jurov@uoregon.edu)[1] **& Thomas Schatz (thomas.schatz@lis-lab.fr)**[2]
[1]Department of Linguistics, University of Oregon, Eugene, OR, USA
[2]Aix Marseille Univ, CNRS, LIS, Marseille, France

Abstract

The human ability to track sequential regularities in speech has been thoroughly studied through both human experiments and computational modeling. These two approaches usually complement each other, in a virtuous cycle between theory and experiment. Yet, they currently seem disconnected when it comes to taking into account the acoustic details of speech. Advances in AI technology now enable computational modeling of statistical learning directly from raw unlabeled continuous recordings. However, when designing and interpreting results from human experiments, speech typically remains modeled in terms of abstract categories – like syllables – without consideration of acoustic details. Here, we bridge this gap by showing that learning from low-level auditory speech representations, rather than from more abstract alternatives, better matches human behavior in statistical learning experiments. This calls into question common assumptions about human learners' representations and underlines the importance of controlling for low-level auditory confounds in the design of statistical learning experiments.

Keywords: statistical learning; cognitive model; speech; acoustics; representations; recurrent neural network

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*