Adaptive speech production: **Targeted millisecond changes in pronunciation as a function of miscommunication**

T. Florian Jaeger, Brain and Cognitive Sciences, University of Rochester
https://www.hlp.rochester.edu/

During typical conversations, human speech transmits about 3-5 syllables every second—coordinating the encoding of meaning into linguistic structures and articulatory motor movements in split seconds. This includes implicit decisions between, alternative sentence structures, words, all the way down to decisions about speech rate and clarity of pronunciations. The inherent complexity of these encoding processes has raised questions about the extent to which the systems underlying language production can possibly be designed to facilitate communication. In this presentation, I approach this question by investigating the consequences of miscommunication. Specifically, I present a series of experiments that investigate how speakers adapt their pronunciations after having been misunderstood (Buz, Tanenhaus, & Jaeger, 2016 and unpublished follow-up work).

Two extreme views can be contrasted. Under the 'egocentric' view, language production is too demanding to make audience design feasible—i.e., language production is primarily concerned with the retrieval of a well-formed structure that encodes the intended meaning. Under the 'communicative efficiency' view, language production trades-off production effort (e.g., the invested attentional resources and articulatory effort) against the expected probability of successful message transmission.

The experiment I discuss investigates articulation of stop-voicing (e.g., "pin") depending on whether minimal pair neighbors (e.g., "bin") are contextually co-present or not, and depending on whether the speaker has previously been misunderstood on similar trials. The findings suggest that co-presence of minimal pair neighbors leads to hyper-articulation, and that this hyper-articulation is further enhanced each time the speaker is misunderstood. Additional analyses show that the resulting hyper-articulation is targeted and contrastive—i.e., targeting the phonetic features that seem to be linked to the misunderstanding, and increasing the probability that interlocutors correctly recognize the word (confirmed by ideal observer analyses). This suggest that talkers can infer the causes of previous miscommunications, and adapt future productions in similar context, so as to enhance to probability of successful message transmission (as predicted by forward models for language production, Jaeger & Ferreira, 2013; Kurumada & Jaeger, 2015; Ferreira, 2019).

If there is time and interest, I will discuss evidence for similarly targeted changes in *grammatical* encoding (production of optional case-marking) from miniature language learning studies.