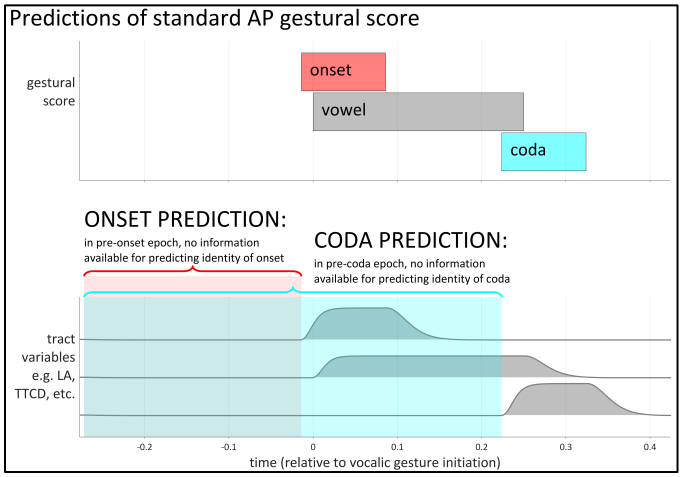
The flow of speech: how Articulatory Phonology under-generates

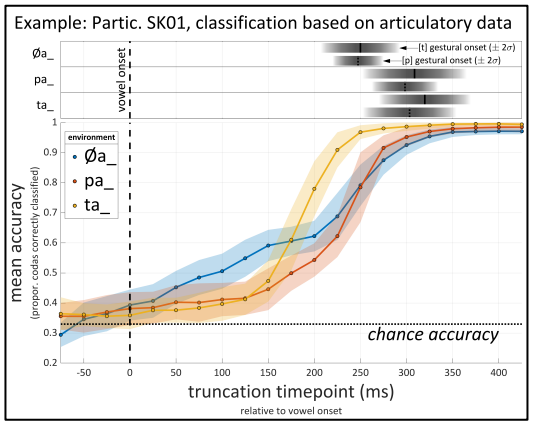
Sam Tilsen

This paper (i) addresses an under-appreciated problem in the theoretical framework of Articulatory Phonology (AP), and (ii) questions the usefulness of conventional event-based analyses of speech data. The main finding is that there exists coarticulatory and anticipatory information in acoustic/articulatory signals that standard AP fails to generate. For example, given a gestural score representation of a CVC syllable as in the figure below, the AP gestural score predicts the following: (1) ONSET PREDICTION: prior to the initiation of gestures associated with an onset consonant, there should be no information in articulatory or acoustic signals that can be used for predicting its identity; (2) CODA PREDICTION: prior to the initiation of gestures associated with a coda consonant, there should be no information that can be used for predicting its identity. To test these predictions, a novel method was applied in which the high-dimensional flow of articulatory and acoustic information was used to train deep neural networks to classify onsets and codas. The performance of these networks on test datasets shows that the onset and coda predictions are not supported.



Method: Acoustic and articulatory (EMA) data were collected from five speakers. Participants produced C/a/C syllables with /p/, /t/, and Ø onsets and codas, resulting in a total of 576 tokens per speaker (9 syllables × 64 repetitions per syllable). On each trial, responses were cued 2000 ms before a go-signal, and participants maintained a prolonged /i/ vowel prior to the response.

Analysis: Typical analyses of speech identify events in a data stream and conduct low-dimensional statistical analyses. Here a high-dimensional approach was employed in which biLSTM networks were trained with articulatory data (EMA sensor positions/velocities) and/or acoustic data (MFCC matrices) to classify onset/coda consonants (see p.2 Fig: *Analysis method*). To test for the presence of coarticulatory and/or anticipatory information in pre-onset or pre-coda epochs, training and evaluation were repeated multiple times with randomly selected training and test subsets, which were systematically truncated to restrict the time-periods from which information was available.



Results: For all participants there was information in pre-onset and pre-coda epochs that resulted in better-than-chance prediction of the onset/coda. The figure on the left shows coda classification accuracy for one speaker in three environments; observe that above-chance accuracy is achieved even when signals were truncated well before the initiation of the coda. In general, it was found that (i) articulatory data allowed for earlier prediction than acoustic data, and (ii) the presence of an onset consonant delayed successful prediction of the coda (see p.2 Fig: *Result details*). The results show that AP gestural scores under-generate with regard to the presence of coarticulatory and anticipatory information. The success of the high-dimensional analysis method raises broader questions about the usefulness of discrete event-based analyses of articulatory control, and the findings speak to the necessity of extending AP to better model anticipatory and coarticulatory phenomena.

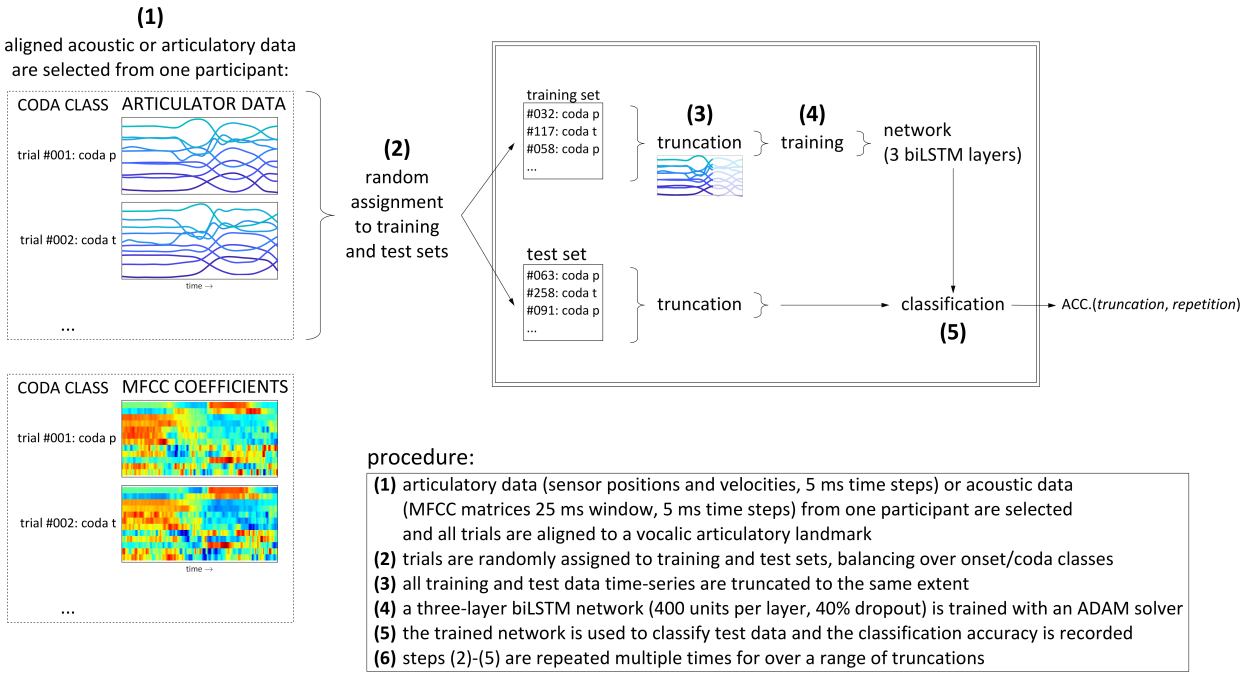


Fig. *Analysis method*. The analysis procedure is described in steps (1)-(6).

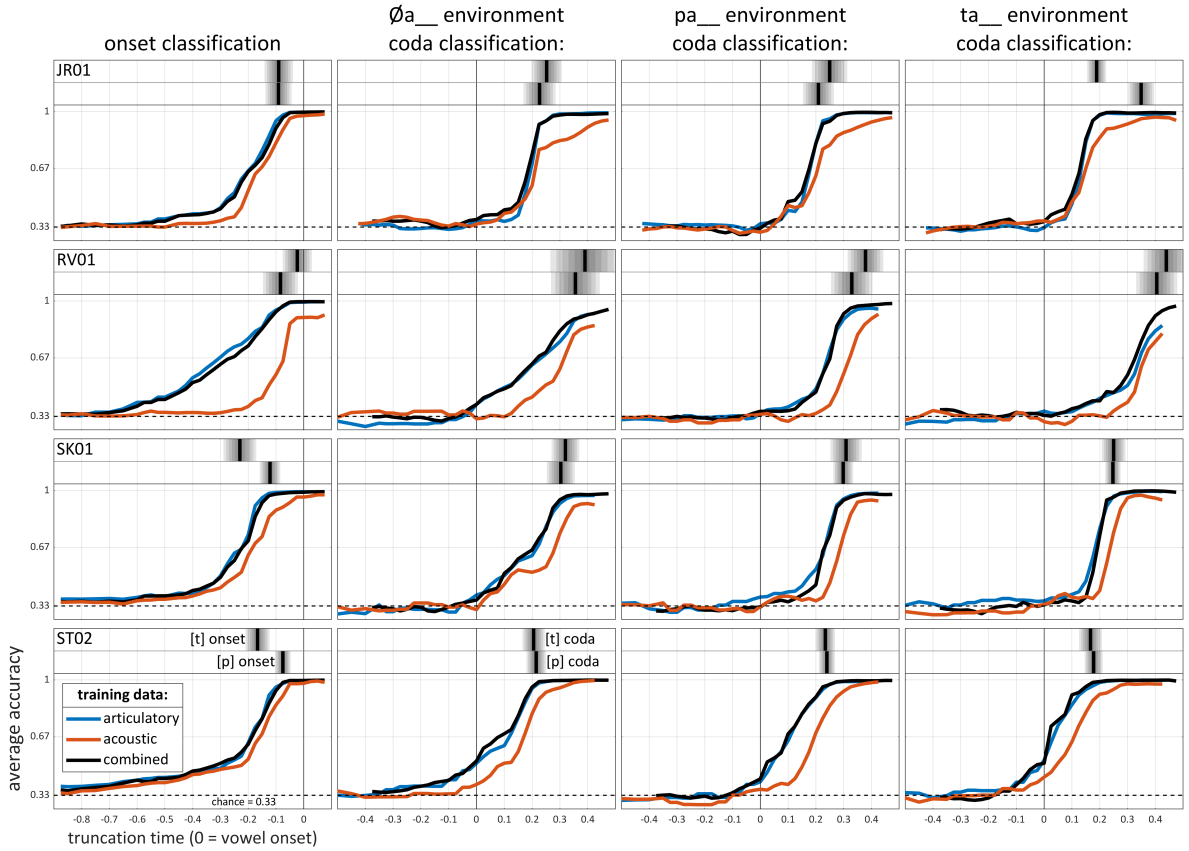


Fig. *Result details*. For four participants (rows), classification accuracy is shown as a function of data truncation for onset classification and coda classification in three environments (columns).