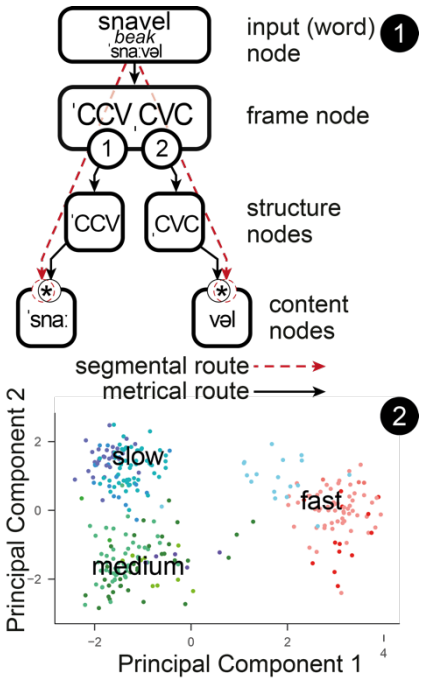
RUN-SPEAKING? SIMULATIONS OF RATE CONTROL IN SPEECH PRODUCTION

Joe Rodd*[1 2] Hans Rutger Bosker[1 3] Mirjam Ernestus[2 1] Antje S. Meyer[1 3] Louis ten Bosch[2 1]
[1] *Max Planck Institute for Psycholinguistics;* [2] *Radboud University, Centre for Language Studies;* [3] *Radboud University, Donders Institute for Brain and Behaviour* * joe.rodd@mpi.nl

That speakers can vary their speaking rate is evident, but how they accomplish this has hardly been studied. Consider this analogy: when walking, speed can be continuously increased, within limits. However, to speed up further, humans must run. Are there multiple qualitatively distinct 'gaits' in speech planning that resemble walking and running gaits? Or is control achieved by continuous modulation of a single gait? This study investigates these possibilities through simulations of a novel connectionist computational model of the cognitive process of speech production, which mimics the temporal characteristics of observed speech.



CONNECTIONIST MODEL Our model, illustrated in Figure 1, is derived from Dell, Burger and Svec's (1997, *Psychol. Rev.* 104(1), 123) model of serial order in language, and sequentially retrieves syllable-level motor plans in response to activation in a word level input node. A frame node mediates, encoding metrical structure and enforcing serial order. This model is the first of its type to predict the precise timing of motor plans and account for the ability to control rate in speech production.

The model has many parameters (connection weightings, thresholds, etc.) that can be adjusted to achieve a specific speaking rate. Different 'regimes' (combinations of parameter settings) can be engaged to achieve different speaking rates. We consider each parameter as a dimension of a high-dimensional 'regime space', in which the regimes occupy different locations.

MODEL TRAINING Our model approximated the distributions of observed syllable durations and syllable overlap durations in the PiNCeR corpus of Dutch disyllabic words produced at fast, medium and slow speaking rates. Syllable onset and offset were identified from the acoustic signal on the basis of spectral instability as an index of syllable overlap. Together, these duration distributions form a 'fingerprint' of the speech production system operating at a given rate. The model was trained separately for each speaking rate, by the natural selection-inspired optimisation algorithm NSGA-III. The training identified parameter values that caused the model to best approximate the 'fingerprint' distributions of each speaking rate in the corpus. The fit of the model was assessed by calculating the Kullback-Leibler divergence between the model's predicted distributions and those taken from the corpus for each speaking rate.

PREDICTIONS In a one gait system, where we 'speed-walk' to speak faster, the regimes used to achieve fast and slow speech are qualitatively similar, but quantitatively different. In regime space, they would be arranged along a straight line. Different points along this axis correspond to different speaking rates. In a multiple gait system, where we 'walk-speak' for slower speaking rates, but 'run-speak' to speak faster, this linearity would be missing. Instead, the arrangement of the regimes would be more disperse, with no obvious relationship between the regions associated with each gait, and an abrupt shift in parameter values to move from speeds associated with 'walk-speaking' to 'run-speaking'.

RESULTS Our model achieved good fits in all three speaking rates. In regime space, the broad arrangement of the parameter settings selected for the different speaking rates is clearly non-axial, suggesting that 'gaits' may be present in the speech planning system (see principal component projection in Figure 2). Thus, we provide the first computationally explicit connectionist account of the ability to modulate the speech production system to achieve different speaking styles.