To speed up, turn up the gain: acoustic evidence of a 'gain-strategy' for speech planning in accelerated and decelerated speech

Joe Rodd[1 2], Hans Rutger Bosker[1], Mirjam Ernestus[2 1], Louis ten Bosch[2 1], Antje S. Meyer[1]

[1] *Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands*

[2] *Radboud University, Centre for Language Studies, Nijmegen, the Netherlands*

That speakers can vary their speaking rate is evident, but how they accomplish this has hardly been studied. The effortful experience of deviating from one's preferred speaking rate might result from the invocation of executive control (EC) processes to modulate the formulation phase of speech planning, namely lexical selection and phonological encoding.

Since there is no existing single model of the entire speech production chain from concept to articulatory movement, we sketch a working model which highlights the distinction between the formulation phase (characterised by retrieval of representations by competitive selection, e.g. [1]–[4]) and the motor execution phase (characterised by direct mapping from planning representations to motoric commands, and crucially, the absence of competition, e.g. [5]). Such a model entails that the broad temporal structure of speech reflects the temporal dynamics of formulation. From this model, we derive two strategies that speakers might invoke to control the formulation network, and thereby their speaking rate: the *gain* strategy, where input activation levels to the formulation network are modulated; and the *threshold* strategy, where selection thresholds are adjusted within the formulation network.

Either strategy can result in earlier or later selection decisions; for example, to speed up, evidence accumulates faster (gain strategy) or the lower threshold is reached earlier (threshold strategy). This results in modulated delay between syllable onsets as speaking rate varies. However, only the gain strategy results in modulated gesture durations as speaking rate varies. This is because only the gain strategy modulates the activation level of the gesture score, which influences the speed at which the gesture score is reproduced. By contrast, the threshold strategy predicts stable gesture durations but modulated overlap between gestures. These predictions are illustrated in Figure 1, panel A.

We present evidence from a picture naming task in Dutch in which 12 participants named pre-familiarised ˈ(C)CV.CVC words (e.g. *snavel* [ˈsnaː.vəl] "beak") from line drawings displayed in groups of 8 arranged on a 'clock face'. A cursor moved clockwise from picture to picture to indicate at which of three rates (132 words per minute, 93 wpm and 66 wpm) participants were required to name the pictures. Annotation was bootstrapped using MAUS forced alignment [6] and manually revised where necessary to yield accurate word onsets and offsets. There were on average 3,754 usable word-tokens for each rate after annotation. To detect regions of acoustically-evident gestural overlap, a novel procedure was employed, detecting excursions of above-average instability of the MFCC vector (c.f. [7]) falling between the MAUS-aligned vowel and consonant centres. This approach was licensed by careful control of segmental content in the target words to maximise correspondence between acoustics and articulation.

From this metric and manually corrected word onsets and offsets, three dependent measures were derived: (1) the duration of the overlap between the syllables; (2) the duration of the first syllable, from word onset to overlap offset; and (3) the duration of the second syllable, from overlap onset to word offset. Mixed effects modelling revealed significant gesture duration modulation (consistent with the gain strategy), but also significant modulation of overlap duration (consistent with the threshold strategy). An examination of effect sizes revealed that the effect on the overlap duration was much smaller than the effects on gesture durations (see Figure 1, panel B). These effect size findings lead us to conclude that speaking rate control in the case of the production of single word utterances is primarily achieved by controlling the activation levels in the formulation network (*gain* strategy), with a subsidiary role for selection threshold manipulation (*threshold* strategy).

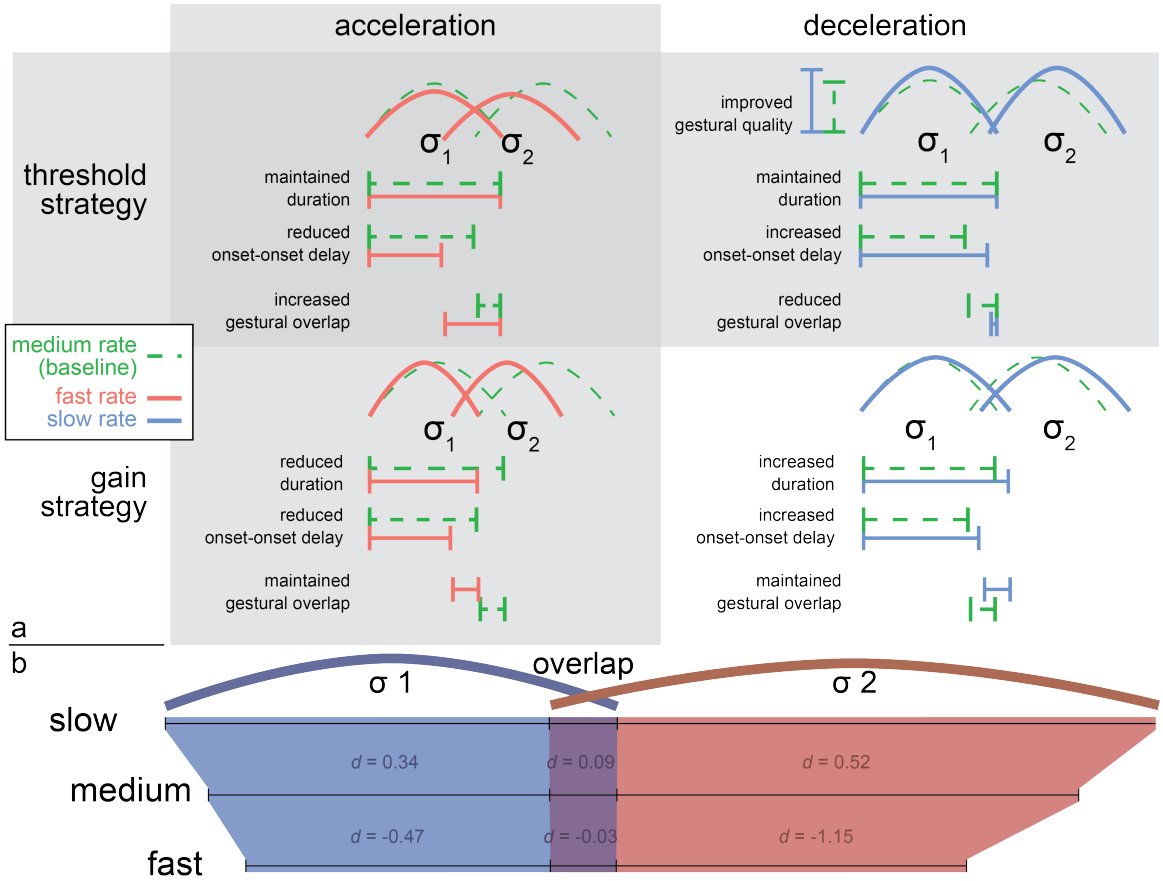

Figure 1. Panel a: *The predictions of the threshold and gain strategies for fast speech and slow speech, presented diagrammatically. The threshold strategy (top) predicts maintained gesture duration, and as a consequence modulated gestural overlap. The gain strategy predicts modulated gestural duration, allowing gestural overlap to be maintained.* Panel b: *the mixed effect model fits presented diagrammatically to illustrate the relative effect sizes in gesture duration modulation and in overlap modulation. σ_1 and σ_2 indicate the first and second syllable, respectively.*

[1] W. J. M. Levelt, A. Roelofs, and A. S. Meyer, 'A theory of lexical access in speech production', *Behav. Brain Sci.*, vol. 22, no. 1, pp. 1–38, 1999.

[2] G. S. Dell, L. K. Burger, and W. R. Svec, 'Language production and serial order: A functional analysis and a model.', *Psychol. Rev.*, vol. 104, no. 1, p. 123, 1997.

[3] G. S. Dell and P. G. O'Seaghdha, 'Stages of lexical access in language production', *Cognition*, vol. 42, no. 1–3, pp. 287–314, 1992.

[4] J. P. Stemberger, 'An Interactive Activation Model of Language Production', in *Progress in the Psychology of Language*, vol. 1, 3 vols, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1985.

[5] J. A. Tourville and F. H. Guenther, 'The DIVA model: A neural theory of speech acquisition and production', *Lang. Cogn. Process.*, vol. 26, no. 7, pp. 952–981, Aug. 2011.

[6] F. Schiel, 'A statistical model for predicting pronunciation', in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 2015.

[7] D.-T. Hoang and H.-C. Wang, 'Blind phone segmentation based on spectral change detection using Legendre polynomial approximation', *J. Acoust. Soc. Am.*, vol. 137, no. 2, pp. 797–805, 2015.